

NBER WORKING PAPER SERIES

SELECTION AND INCENTIVE EFFECTS OF ELECTIONS:
EVIDENCE FROM STATE SUPREME COURTS

Elliott Ash
W. Bentley MacLeod

Working Paper 22071
<http://www.nber.org/papers/w22071>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2016, Revised May 2019

Previously circulated as "The Performance of Elected Officials: Evidence from State Supreme Courts." We thank Yisehak Abraham, Ankeet Ball, Josh Brown, Josh Burton, Matthew Buck, Eamonn Campbell, Zoey Chopra, Daniel Deibler, Seth Fromer, Gohar Harutyunyan, Archan Hazra, Montague Hung, Dong Hyeun, Mithun Kamath, James Kim, Michael Kurish, Jennifer Kutsunai, Steven Lau, Sharon Liao, Sarah MacDougall, Sam Meshoyrer, Justin McNamee, Sourabh Mishra, Brendan Moore, Arielle Napoli, Karen Orchansky, Bryn Paslawski, Olga Peshko, Quinton Robbins, Ricardo Rogriguez, Jerry Shi, Shawn Shi, Carol Shou, Alex Swift, Holly Toczek, Tom Verderame, Sam Waters, Sophie Wilkowske, John Yang, Geoffrey Zee, Fred Zhu, and Jon Zytnick for their meticulous help in assembling data and other research assistance. We thank Daniel Chen, Tom Clark, John Ferejohn, Sanford Gordon, Chris Hanretty, Jon Kastellec, Lewis Kornhauser, Eric Posner and the participants at Princeton University Conference on Bureaucrats, SIOE meetings at Harvard Law School, NYU Law and Economics Workshop, and Conference on Empirical Legal Studies in Europe for helpful comments. Columbia University's Program for Economic Research, Columbia Law School, and the National Science Foundation Grant SES-1260875 provided financial support for this research. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2016 by Elliott Ash and W. Bentley MacLeod. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Selection and Incentive Effects of Elections: Evidence from State Supreme Courts
Elliott Ash and W. Bentley MacLeod
NBER Working Paper No. 22071
March 2016, Revised May 2019
JEL No. J24,K4

ABSTRACT

Using data on state supreme court judges for the years 1947 through 1994, we find that judges selected by nonpartisan elections and judges selected by technocratic merit commissions produce higher-quality work than judges selected by partisan elections. Election-year pressure reduces work output, but only partisan elections reduce work quality. Moving from nonpartisan elections to uncontested elections increases work quality for incumbent judges, while there is no effect on incumbent judge performance when moving from partisan to nonpartisan elections, or when moving from partisan to uncontested elections, consistent with the hypothesis that non-partisan judges have a greater intrinsic value for quality.

Elliott Ash
ETH Zurich
IFW E47.1
Zurich 8044
Switzerland
ashe@ethz.ch

W. Bentley MacLeod
Department of Economics
Columbia University
420 West 118th Street, MC 3308
New York, NY 10027
and NBER
bentley.macleod@columbia.edu

The aim of every political constitution is, or ought to be, first to obtain for rulers men who possess most wisdom to discern, and most virtue to pursue, the common good of the society; and in the next place, to take the most effectual precautions for keeping them virtuous whilst they continue to hold their public trust.

– *Federalist No. 57*

1 Introduction

The decisions of public officials have a large impact upon our daily lives, yet they typically face weak incentives for good performance. In the case of judges, weak incentives are a design feature justified by the need to have them issue unbiased decisions in the public interest.¹ This is particularly important for appellate courts in common-law systems, such as the United States, where judicial decisions have the power of law.²

In the absence of explicit pecuniary rewards, what motivates good performance for public officials? One possibility is the intrinsic reward from doing a good job, which may arise when organizations have a clear “mission” that can motivate their members (Wilson, 1989). Dewatripont et al. (1999) study the design of rewards in the presence of mission, and in particular highlight the role of “professionalism” in encouraging task-specific skills, which in turn increase the professional’s intrinsic incentive for work (see White (1959) and Wilensky (1964)).³

These ideas highlight the importance of selecting the right type of public official who values work quality, and providing incentives that do not interfere with these values. Alesina and Tabellini (2007, 2008) analyze decision quality under elections or bureaucracy. Tenured “bureaucrats” are preferred for technical tasks, where organizations can evolve a mission, and for protecting the rights of minorities. In contrast, elected “politicians” are more sensitive to outcomes, and to the preferences of the median voter. The choice of one or the other system is a practical question that depends upon the empirical magnitude of these factors.⁴

¹See for example Epstein et al. (2013). The optimality of this policy follows directly from the Holmström and Milgrom (1991) multi-tasking model when there are no good measures of decision quality.

²The state court websites provide up-to-date information on how appellate courts work in each state. See <https://www.ohioabar.org/forpublic/resources/lawyocanuse/pages/lawyocanuse-448.aspx> for a concise explanation. See Landes and Posner (1980) on the role of judges in common law systems. The work of Djankov et al (2008) has been influential in documenting how variation in legal systems, including the effectiveness of judicial decision making, can affect economic performance. Klerman and Mahoney (2005) provide additional historical evidence linking increases in judicial independence with increased economic performance.

³Francois (2000) uses this idea to build a theory of non-profit firms.

⁴Other presentations of the political agency model (focusing on selection and incentive effects of elections) include Fearon (1999), Besley (2006), and Alt et al. (2011). Maskin and Tirole (2004) consider the optimal

The purpose of this paper is to estimate the effect upon performance of changing the appointment procedure for state appellate courts. We do this by exploiting the changes in the way state supreme court judges are selected in the United States over the period 1947-1994. These courts serve as the state judiciary’s analogue to the U.S. Supreme Court and have the authority to review laws produced by state legislatures and decisions produced by lower state courts. Thus state supreme court judges are some of the most powerful officials in state government. These judges are the last appeal on some of the most important features of common law, including most rules on contract, property, tort, and crimes.

The identification of the effect of appointment procedure relies on two ingredients. First, over this period of time U.S. states experimented with different methods of appointing judges. Second, the job of a supreme court judge does not change much over the course of the career, and it does not vary across states. The main component of the job is to write an opinion explaining the court’s decision. This involves working (mainly with clerks, and not with other judges) to research the relevant case law, reason through the implications for present and future litigants, and introducing a precedent that future judges in the state must follow.⁵ Judges do not write all the text of their opinions, however they are responsible for each decision they sign. The quality of judges depends not only upon their native skills as a writer of legal arguments, but also as an individual who hires and manages law clerks that assist in this activity. Thus, if we find that some judges perform better than others, then we are providing some direct evidence on the importance of management for performance.⁶

We construct two classes of performance measures. The first of these is “output”. This includes number of cases heard, total words written, and citations to other cases. These roughly measure time on the job. The second class of performance measures correspond to “quality”. Given that the job of an appellate judge is to clarify and in some cases create new law, legal scholars measure quality by the frequency with which the decisions of a judge are cited in future cases, both within-state and out-of-state.⁷

While the task of a judge is similar from state to state, the rules for selecting and retaining appellate judges vary across states and over time. These rules are listed in Table 1, with rule changes indicated by cell borders. A graphical representation of these changes is included in

choice of institution (removable “politician” or unaccountable “judge”); election pressure can cause officials to modify their decisions to reflect the interests of the electorate.

⁵See Hanssen (e.g. 1999); Helland and Tabarrok (e.g. 2002); Hanssen (e.g. 2004); Hall and Bonneau (e.g. 2006); Hall (e.g. 2007); Kritzer (e.g. 2011, 2015) for details on the operation of these courts.

⁶Bloom and Reenen (2007) document that managers do matter for the performance of private firms.

⁷See Posner (2008) for a seminal discussion of the work of judges and what we mean by quality. Choi et al. (2010) discuss in detail why citations are used to measure judicial performance.

Appendix Figure 3.

In this study we are concerned with three types of appointment systems: partisan elections, non-partisan elections and appointment by the state governor. The first system, partisan elections, is used for both selection of new judges and retention of incumbent judges. For these elections, judges represent a political party, Republican or Democrat, that is clearly identified on the ballot. They must win a primary election for their party before running in a general election. Incumbent judges rarely face a credible challenge in the primary, but in the general election they usually face a challenger from the opposing political party.

The second system, nonpartisan elections, are also used for both selection and retention. In this system there are competitive elections, but there are no primaries and party affiliations are not on the ballot. There are generally two candidates, an incumbent and a challenger, but the incumbent is not identified as such.

The third major system is merit selection with uncontested retention elections, also known as the Missouri Plan. In this system, judges are nominated by a commission of experts – senior attorneys and retired judges – and confirmed by the governor. Incumbent judges face an up-or-down retention vote with no challenger. This system is designed to be more meritocratic, and to impose weaker political incentives, than electoral selection.⁸

There is an extensive discussion in the political science literature of the reasons for these different systems, and the effect they have upon judicial decisions, that we discuss briefly in the next section. This work is concerned with the politics of decision-making, and when judges make decisions that pander to one political party or the other. Our concern here is with the labor economics of this particular market for experts. By an expert we mean any person who is engaged to carry out tasks for which the employer is not qualified to do. Doctors, lawyers or financial advisors are all experts whose demand for services depends upon consumers buying (or in our case voting) with their choices. The question is whether or not the non-expert decision maker is able to distinguish the quality of the expert.

In our context, the question we are interested in is whether the appointment system positively selects judges for their ability. In our earlier work, Ash and MacLeod (2015), we

⁸There is a fourth hybrid system, where judges are initially selected through partisan elections but thereafter face uncontested retention elections. California has governor appointment but uncontested retention elections. The other states either have some combination of governor or legislative appointment, both for initial selection and for period retention. In Massachusetts, New Hampshire, New Jersey, and Rhode Island, judges have lifelong tenure. In Ohio and Michigan, judicial elections are difficult to classify within the partisan/nonpartisan dichotomy because they have partisan primaries and nomination processes, but the political party is not on the ballot in general elections. Following Nelson et al. (2013), we classify these states as partisan selection and nonpartisan retention. Alternative codings, or leaving them out of the analysis, does not change our results.

Table 1: Judicial Selection and Retention Systems

State (Years)	Selection	Retention	State (Years)	Selection	Retention
Alaska	Merit	Uncontested	New Hampshire	Governor	Tenure
Alabama	Partisan	Partisan	North Carolina	Partisan	Partisan
Arkansas	Partisan	Partisan	North Dakota	Non-Partisan	Non-Partisan
Arizona (-1974)	Non-Partisan	Non-Partisan	Nebraska (-1962)	Partisan	Partisan
Arizona (1975-)	Merit	Uncontested	Nebraska (1963-)	Merit	Uncontested
California	Governor	Uncontested	New Jersey	Governor	Tenure
Colorado (-1966)	Partisan	Partisan	New Mexico (-1988)	Partisan	Partisan
Colorado (1967-)	Merit	Uncontested	New Mexico (1989-)	Partisan	Uncontested
Connecticut	Governor	Governor	Nevada	Non-Partisan	Non-Partisan
Delaware	Governor	Governor	New York (-1976)	Partisan	Partisan
Florida (-1971)	Partisan	Partisan	New York (1977-)	Governor	Governor
Florida (1972-1976)	Non-Partisan	Non-Partisan	Ohio	Partisan	Non-Partisan
Florida (1977-)	Merit	Uncontested	Oklahoma (-1967)	Partisan	Partisan
Georgia (-1984)	Partisan	Partisan	Oklahoma (1968-)	Merit	Uncontested
Georgia (1985-)	Non-Partisan	Non-Partisan	Oregon	Non-Partisan	Non-Partisan
Iowa (-1962)	Partisan	Partisan	Pennsylvania (-1968)	Partisan	No Retention
Iowa (1963-)	Merit	Uncontested	Pennsylvania (1969-)	Partisan	Uncontested
Idaho	Non-Partisan	Non-Partisan	Rhode Island	Governor	Tenure
Illinois (-1964)	Partisan	Partisan	South Carolina	Legislature	Legislature
Illinois (1965-)	Partisan	Uncontested	South Dakota (-1980)	Non-Partisan	Non-Partisan
Indiana (-1970)	Partisan	Partisan	South Dakota (1981-)	Merit	Uncontested
Indiana (1971-)	Merit	Uncontested	Tennessee (-1971)	Partisan	Partisan
Kansas (-1958)	Partisan	Partisan	Tennessee (1972-1977)	Merit	Uncontested
Kansas (1959-)	Merit	Uncontested	Tennessee (1978-)	Partisan	Partisan
Kentucky (-1975)	Partisan	Partisan	Texas	Partisan	Partisan
Kentucky (1976-)	Non-Partisan	Non-Partisan	Utah (-1951)	Partisan	Partisan
Louisiana	Partisan	Partisan	Utah (1952-1985)	Non-Partisan	Non-Partisan
Maine	Governor	Governor	Utah (1986-)	Merit	Uncontested
Maryland (-1976)	Non-Partisan	Non-Partisan	Vermont (-1971)	Legislature	Legislature
Maryland (1977-)	Merit	Uncontested	Vermont (1972-)	Governor	Legislature
Massachusetts	Governor	Tenure	Virginia	Legislature	Legislature
Michigan	Partisan	Non-Partisan	Washington	Non-Partisan	Non-Partisan
Minnesota	Non-Partisan	Non-Partisan	Wisconsin	Non-Partisan	Non-Partisan
Missouri	Merit	Uncontested	West Virginia	Partisan	Partisan
Mississippi	Partisan	Partisan	Wyoming (-1972)	Non-Partisan	Non-Partisan
Montana	Non-Partisan	Non-Partisan	Wyoming (1973-)	Merit	Uncontested

Notes. This table lists the elections systems for state supreme court judges observed in our data. Election-system reforms indicated by cell borders.

found that on average appellate court judges do have a taste for quality. We were able to show this by exploiting changes in their workload, that in turn led them to write longer and more influential opinions. In the case of elections, we cannot expect voters on average to be well informed regarding the quality of a judge. Moreover, since we do not have data on the output of all potential candidates, we cannot measure the quality difference between winners and losers. Rather, our approach is to ask the practical question of whether a state's choice to change from, say, a partisan election system to a non-partisan election system leads to a change in the quality of the court.

Choi et al. (2010) look at this question using data on the opinions written by all state appellate court judges over a period of three years (1998-2000). They find that in the cross section states with appointed judges have higher quality decisions, but produce fewer decisions. These results are consistent with our results in Ash and MacLeod (2015) suggesting that judges face a trade-off between quality and output. Given that these results are from a cross-section they do not answer the question of what will happen when a given state changes their appointment system.

We have constructed a data set of all state supreme court decisions from 1947 until 1994 that is matched to a number of judge characteristics and state rules on judicial appointment. The data on decisions was provided by Bloomberg Law, which includes the number of (positive) citations to each decision. With this data we can construct a number of measures of performance and output. We find that states that move from partisan elections to either nonpartisan elections and or technocratic merit commissions choose judges who produce higher-quality work. We also find evidence that judges do care about re-election and put effort into gaining re-election. This follows from the observation that in an election-year a judge's work output falls. Interestingly, if a judge is up for re-election under a partisan system we also observe a fall in work quality. Moving from nonpartisan elections to uncontested re-elections increases work quality for incumbent judges, while there is no effect on incumbent judge performance when moving from a partisan to a nonpartisan re-election system, or when moving from partisan to uncontested re-election. These results are consistent with the hypothesis that partisanship affects judge performance through the selection of judges that have a lower intrinsic value for quality.

The agenda of the paper is as follows. In the next section we briefly review the relevant literature. Section 3 discusses the role of information in the judicial selection process. The data and performance measures are discussed in Section 4, followed by a discussion of our identification strategy and empirical analysis. Section 6 concludes the paper. The Appendix

contains results from a number of robustness checks.

2 Literature

The literature can be roughly divided into two distinct questions. In economics there is a literature that is concerned with understanding the quality of civil servants, and how different appointment systems affect this quality. In our case these are judges, but similar issues apply to any individual who is chosen to make complex decisions on our behalf. For the most part voters do not have the skill set to carry out the choices that the appointed individual is carrying out, yet they must decide who they should support for appointment.

There is a literature in economics showing that electoral pressures do affect decision making. Besley and Case (1995) use the variation in term limits to show that state governors do respond to electoral pressure and reduce state spending where there is a binding term limit. Besley and Coate (2003) show that electricity prices are lower in states with elected (versus appointed) utility regulators, consistent with elected regulators being more responsive to consumers. List and Sturm (2006) show that electoral incentives drive decision-making by governors on environmental issues. Ferraz and Finan (2008) find that voters respond to information on corruption, and therefore serve to select out the worst politicians. Ferraz and Finan (2011) find that mayors who are term-limited (and therefore face no re-election incentives) are more corruption, consistent with an incentive effect against corruption. Pande (2011) reviews the literature in development economics on political accountability and finds that voters respond to information about the political process and that politician performance improves electoral accountability. In particular, the quality of information regarding candidates plays a central role in explaining the quality of elected politicians.

Like Choi et al. (2010) discussed above, Lim and Snyder (2015a) use the cross-section variation in election systems to study the relationship with various measures of trial judge quality in a sample of 39 states from 1990 - 2010. They find that voters are strongly influenced by party cues in partisan elections, but add the caveat that trial court judges are less important than appellate court judges in making law, which might result in less voter scrutiny.

There is also an important and voluminous literature documenting the significant impacts of political motivations on decision-making. Hall and Bonneau (2006) have an interesting paper that looks at the effect of a challenger upon the success of an incumbent judge using data from 21 states from 1990 to 2000. They find that voters do respond to challenger

quality. There are also a number of papers that focus upon the politics of decision making. Specifically, does the appointment system affect the *values* that use to evaluate cases. A key finding is that stronger electoral pressures are associated with harsher criminal sentences (Huber and Gordon, 2004; Gordon and Huber, 2007; Lim, 2013; Berdejo and Yuchtman, 2013; Iaryczower et al., 2013; Canes-Wrone et al., 2014; Park, 2014; Lim et al., 2015). Looking at federal judges, Epstein et al. (2013) review the evidence that decisions tend to reflect the ideological leanings of the president that appointed them. Besley and Payne (2013) find that elected judges are more likely to support anti-discrimination law. Shepherd (2009) and Canes-Wrone et al. (2010) find that retention elections cause state appellate judges to vary their decisions in subsets of politically sensitive cases. Fournaies and Hall (2018) show that U.S. state legislators who are term-limited decrease effort, in terms of sponsoring bills, acting on committees, and being present for floor votes.

Finally, there is a literature that looks at the information processing by voters. Early work by Rahn (1993) shows that voters use partisan identifiers as a decision making short cut. Using a combination of survey and experimental data, Bonneau and Cann (2015) and Lim and Snyder (2015a) also find that partisan identification plays an important role in voter decision making. Recently, A. Kirkland and Coppock (2017) provide some direct experimental evidence that if voters cannot observe partisan labels then they put more weight upon observable judge characteristics that reflect quality. In the next section we discuss how such a mechanism may lead to the quality of judges varying with the type of system used to appoint them.

3 Mechanisms

In our data we observe the change in appointment system and link these changes to measures of judge performance. The literature shows that when moving from a partisan to non-partisan system voters are likely to use more information regarding judge ability. In this section we show how formally this may lead to variation in the quality of selected judges. The theoretical literature on voting in political science is well-developed and sophisticated (see Ashworth and de Mesquita (2008) and Ashworth et al. (2017) for comprehensive reviews of this literature). However, empirically we only look at the most basic features of the environment - the division of appointment systems into three crude categories, and the effects that these have upon basic performance measures.

The relevant question is: How do these appointment systems change the performance of

judges? To answer this question, suppose voters are faced with the choice between judge A and B . We follow Condorcet (1785) and view voting as an information revelation problem. The voter's choice is based upon some signal s_j of a latent factor, q_j . It is also the case that voters care about the political views of judges, and hence they may be biased for one judge over another. Let b_j denote the bias term. Hence we suppose:

$$s_j = b_j + q_j + \gamma_j, j \in \{A, B\}.$$

It is possible that the election system can affect the pool, both in terms of the levels and variance. One can model this by setting:

$$\begin{aligned} q_A &= m + \delta/2, \\ q_B &= m - \delta/2, \end{aligned}$$

where m is the mean ability in the electoral pool, and δ is a measure of the variance. Note that without loss of generality judge A is assumed to have higher ability. Given that the elections are relative races, then like other parameters, bias can be viewed as a single parameter in favor of Judge A :

$$\begin{aligned} b_A &= b/2, \\ b_B &= -b/2. \end{aligned}$$

This allows for nice closed-form solutions for the expected ability. Moreover, it is possible that the parameters vary with the appointment system and future behavior of the judge. At the time of appointment, the parameters are assumed to take on their expected values.

Under the hypothesis that there is a common latent factor, and that voters prefer higher q to lower q , then we may suppose that there is a representative voter who chooses the judge with the highest value (see Rothstein (1991); Saporiti and Tohmé (2006)). Thus, the judge selected in period τ is given by:

$$j_\tau^* = \operatorname{argmax}_{j \in \{A, B\}} \{s_A, s_B\}. \quad (3.1)$$

The expected quality of the judge chosen in period τ is given by:

$$\hat{q}_\tau = E \{q_{j_\tau^*}\} = m + \delta \left(F \left(\frac{\delta + b}{\sigma} \right) - \frac{1}{2} \right), \quad (3.2)$$

where $\sigma^2 = \sigma_A^2 + \sigma_B^2$ is the variance of $\gamma_B - \gamma_A$, and $F()$ is the cumulative distribution function for the Normal distribution.

3.1 Judicial Effort

Judges have the potential to affect these signals through two avenues. One is to provide more information regarding themselves so that the voters can more accurately assess their quality. Second, they can pander to the voters regarding their political views, and thus affect the bias in their favor. Let the effort of judge j be given by $\vec{e}_j = \{e_j^I, e_j^P\}$, and total effort be given by $\vec{e} = \{\vec{e}_A, \vec{e}_B\}$, where I indicates providing information (not disinformation!) and P refers to pandering.

Suppose that bias is given by:

$$b(\vec{e}^P) = \bar{b} + P(e_A^P) - P(e_B^P),$$

where $P() \geq 0$ is a bounded, differentiable, strictly concave ‘‘pandering’’ function that satisfies $P(0) = 0$. It is also assumed that variance $\sigma_j^2(e_j^I)$ is a concave, differentiable, strictly decreasing function of effort that is bounded below by zero. Let $\sigma(\vec{e}^I) = \sqrt{\sigma_A^2(e_A^I) + \sigma_B^2(e_B^I)}$ be the variance of the difference in error terms as a function of judge investment into information transmission. The cost of effort for Judge j is given by $C(e_j^I + e_j^P)$, where it is assumed that $C'(0) = 0$, $C''(e) > 0$ and there is an \bar{e} such that $\lim_{e \rightarrow \bar{e}} C(e) = \infty$. Suppose that the utility that judges get from obtaining office is $U_j^* > 0$. Then the payoffs of the judges are:

$$\begin{aligned} U_A(\vec{e}) &= U_A^* \times Pr\{s_A > s_B | \vec{e}\} - C(e_A^I + e_A^P), \\ &= U_A^* \times F\left(\frac{\delta + b(\vec{e}^P)}{\sigma}\right) - C(e_A^I + e_A^P). \end{aligned}$$

$$U_B(\vec{e}) = U_B^* \times F\left(\frac{-(\delta + b(\vec{e}^P))}{\sigma}\right) - C(e_B^I + e_B^P).$$

Notice that Judge A has an incentive to invest in providing information if and only if $\delta + b > 0$, while Judge B will invest if and only if $\delta + b < 0$. Thus, only one of the two judges has an incentive to invest in providing information, depending upon whether or not the bias is sufficiently strong to overcome the quality differences.

In contrast, both judges have an incentive to invest in manipulating the bias. Consider the case where the bias is such that Judge A chooses to provide information to voters ($e_A^{I*} > 0$). In that case the first order conditions for an effort equilibrium are given by

$$\frac{\sigma^*}{U_A^* f\left(\frac{\delta + b^*(\bar{e}^P)}{\sigma^*}\right)} \times C'(e_A^{I*} + e_A^{P*}) = \frac{dP(e_A^{P*})}{de_A^P} = \frac{\delta + b^*}{\sigma^*} \frac{\partial \sigma(e_A^{I*})}{\partial e_A^I},$$

$$\frac{\sigma^*}{U_B^* f\left(\frac{-(\delta + b^*(\bar{e}^P))}{\sigma^*}\right)} \times C'(e_B^{P*}) = \frac{dP(e_B^{P*})}{de_B^P}.$$

Notice that the Normal probability density function, $f()$, is symmetric and hence the term in front of costs C' is the same in both expressions. Suppose that the gain from winning is the same for both judges ($U_A^* = U_B^*$), and Judge A chooses $e_A^I > 0$, then it must be the case that the level of pandering by Judge B is more than by Judge A ($e_B^{P*} > e_A^{P*}$). In other words, the less-skilled judge does not invest in providing information on his performance, but panders more to voters than the more highly skilled Judge A. This effect is reversed if the bias is sufficient to make Judge B more likely to win.

3.2 Judicial Selection

Let us now consider how election success is affected by the parameters that are exogenous from the perspective of voters. Suppose that they have rational expectations regarding the parameters of the system. In general we cannot expect voters to have a great deal of high-quality information regarding candidates, though as we discussed above there is evidence that votes do have some information regarding judges. Given that both judges have an incentive to engage in pandering, to a first order we can assume that it cancels out, and thus we can take the baseline level of bias, \bar{b} , as given.

Given this and the equilibrium variance of signal quality derived from the previous section, we have that the average quality of an elected judge in period τ is given by:

$$\hat{q}_\tau = m + \delta \left(F\left(\frac{\delta + \bar{b}}{\sigma}\right) - \frac{1}{2} \right), \quad (3.3)$$

where $\sigma^2 = \sigma_A^2 + \sigma_B^2$ is the variance of the difference in errors. From expression (3.3) we

immediately get:

$$\frac{\partial \hat{q}_\tau}{\partial m} > 0, \tag{3.4}$$

$$\frac{\partial \hat{q}_\tau}{\partial b} > 0, \tag{3.5}$$

namely, an increase in mean quality of the pool increases expected quality, all else held fixed. Given that Judge A is assumed to be better, then bias in favor of Judge A increases the probability that he/she wins, and hence increases average quality. Hence, the effect (3.5) is of practical use if and only if we can tell *ex ante* who is the better judge.

Next, notice that the effect of the two variance measures (δ and σ^2) also depends upon the bias term. The *positive bias* case occurs when $\delta + b > 0$. This ensures that the probability of the better judge (Judge A) winning is greater than 1/2, in which case we have:

$$\frac{\partial \hat{q}_\tau}{\partial \delta} > 0, \tag{3.6}$$

$$\frac{\partial \hat{q}_\tau}{\partial \sigma_j^2} > 0. \tag{3.7}$$

Ex ante, the higher quality judge is not known by the voters, and hence it is more useful to think about bias in terms of its strength. When the size of the bias is less than the variance of the pool, $|b| < |\delta|$, we call this the *weak bias* case, which also implies (3.6 - 3.7).

In the *strong bias* case, the result depends upon the sign. If the bias is in favor of the better candidate, then we get the above comparative statics. However, if $\delta + b < 0$, then we get:

$$\frac{\partial \hat{q}_\tau}{\partial \sigma_j^2} < 0. \tag{3.8}$$

The effect of δ is ambiguous in this case and depends upon the relative magnitudes of the bias and the variance.

3.3 Application to Empirical Context

Choi et al. (2010) begin their article on judicial quality with a quote from Justice O'Connor, concurring in *Republican Party of Minn. v. White*, 536 U.S. 765, 792 (2002):

“If the state has a problem with judicial impartiality, it is largely one the state

brought upon itself by continuing the practice of popularly electing judges.”

Our goal is to explore the validity of this claim by exploring the natural experiment that arises when states change the system they use from appointing judges. Consider first a non-partisan system. In such a system judges are not identified with any political party, and hence a priori the voters may not know where they stand on certain issues that are associated with parties. In the context of our model, this is captured by setting the bias term $\bar{b} = 0$. This also implies that in an election year a non-partisan judge may have a greater incentive to pander to voters than under a partisan system, to let them know where he stands on sensitive issues. Hence, compared to a partisan system, we might expect non-partisan judges to spend more time in election year politics, which in turn may be reflected in less effort on the bench. In our data we can observe election year performance, and we can see how it affects both total output and the quality of decision making. In the case of appointed judges, their reappointment is not determined by voters and hence we would not expect to see any performance fall during years they are reappointed.

Thus, the null hypothesis is that in an election year judges work at the same level as in a non-election year. We expect this null to be rejected in the case of partisan and non-partisan elections, with possibly a larger effect in a non-partisan election.

In terms of the average quality of judges the case is far from clear. As we discussed above, there is evidence to support the hypothesis that voters are influenced by the political party. However, this does not necessarily imply less competition. For example, if party A is more favored in a state, then individuals who want to be judges would know that they should compete in A’s primaries (there are cases of judges changing political affiliation for this purpose). See Hirano and Snyder Jr (2014) for some evidence on this point. In the case of a merit panel we can suppose that they have better quality information. Thus, again the null hypothesis is that the appointment system has no effect upon the quality of judges appointed. The theory suggests two alternatives to the null hypothesis:

1. Suppose that both candidates are drawn from the same distribution. In partisan elections voters have less incentive to gather information relative to the non-partisan system, and hence we expect judge quality to be highest under a merit system, followed by a non-partisan election system, and finally the partisan system would have the lowest quality judges.
2. Suppose that political parties use superior information to choose candidates, and the distribution of judges under a merit system is the same as the distribution of a randomly

chosen judge under the non-partisan system, then the quality of judges is highest under the partisan system, followed by the non-partisan system, then then the merit system would have the lowest quality.

The second result follows from (3.1). Namely, if appointment systems are not competitive in the sense that the distribution of judges they select is the same as a single judge presenting themselves for office, then election systems introduce an element of competition that improves quality. Whether or not a partisan system is superior to a non-partisan system depends upon the quality of the primary system.

From the results in Ash and MacLeod (2015) we know that judges face a trade-off between output and quality. Those results follow from exogenous shocks to workload generated by the addition of intermediate appellate courts, and hence identified separately from the appointment system. Left to their own devices we found that judges, like many academics, prefer quality over quantity. By this, we mean they seem to prefer to write longer opinions that are cited by their colleagues in the future. By measuring how the appointment system affects output and quality separately we should also be able to uncover some information regarding voter preferences over these different performance measures.

4 Data and Performance Measurement

4.1 Data Overview

The data-set used for the empirical analysis is an update and extension of that used in Ash and MacLeod (2015). It merges information on judge biographies, state-level court institutions, and published judicial opinions. For this paper, the dataset was checked and rebuilt, with more information relevant to elections added or updated. These data allow panel estimates on the effects of court institutions on judge performance.

There are 1,628 state supreme court judges in our data. Table 2 reports summary statistics on the characteristics of judges working in one of the three selection systems discussed in the introduction. For many of the variables, the systems are comparable. Relative to the partisan judges, the nonpartisan and merit judges are more likely to be female. Merit judges are the most likely to have judicial experience, while partisan judges are the most likely to have political experience. Nonpartisan and merit judges have longer career lengths. Merit judges are the least likely to lose re-election.

Our performance measures were constructed from published state supreme court opinions

Table 2: Summary Statistics on Judge Characteristics by Selection System

	Partisan Elections		Non-Partisan Elections		Merit Selection	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
<u>Background</u>						
Start Age	53.6969	8.8354	52.8235	8.3454	52.1143	7.8792
Female	0.0305	0.1721	0.0663	0.2491	0.0616	0.2410
Top School	0.0973	0.2966	0.1040	0.3058	0.1081	0.3114
<u>Previous Experience</u>						
Private Practice	0.6862	0.4644	0.8141	0.3897	0.6645	0.4737
Judiciary	0.6082	0.4886	0.5630	0.4969	0.6818	0.4673
Politics	0.2818	0.4503	0.2269	0.4197	0.1118	0.3162
Academia	0.0879	0.2834	0.1076	0.3105	0.1060	0.3088
<u>Partisan Affiliation</u>						
Republican	0.5517	0.4675				
Democrat	0.4483	0.4412				
Career Length	11.8401	8.6620	13.6512	9.6248	13.2120	7.0797
<u>How Ended</u>						
Retired	0.3293	0.4703	0.3896	0.4883	0.3825	0.4871
Resigned	0.1111	0.3145	0.2098	0.4077	0.0553	0.2291
Died in Office	0.1070	0.3094	0.1144	0.3188	0.0553	0.2291
Lost Election	0.0650	0.2468	0.0409	0.1983	0.0138	0.1170
Impeached	0.0054	0.0735	0.0000	0.0000	0.0046	0.0679
Judges	738		367		217	

Notes. Biographical information by judge election system. Observation is a judge. *Start Age* is judge age upon joining the court. *Female* is a dummy for being female. *Top School* means the judge attended law school at Yale, Harvard, Columbia, Stanford, or Chicago. The *Previous Experience* items equal one if the judge has previous experience in the respective area. *Republican* is a dummy for being Republican, *Democrat* for being Democrat. *Career Length* is number of years working on the court, conditional on having left the court before 2014. The *How Ended* items equal one if the judge's state supreme court judgeship ended for this reason.

for the years 1947 through 1994, obtained (along with some annotated meta-data) from bloomberglaw.com. The full sample includes 1,024,261 cases. We drop opinions that do not have a named author (per curiam decisions), resulting in a sample of 404,928 majority opinions. This is an average 47.2 cases per judge per biennium.

4.2 Measuring Judge Performance

An important step in this research is to provide an effective measure of judge performance. We focus on two simple metrics for judge performance, work output and work quality. The measures build off of work by previous researchers, in particular Choi et al. (2010) and Epstein et al. (2013).

The baseline measure of *work output* is the total number of words written by a judge in opinions during a year on the job. This is a measure of the total volume of opinion-writing work that a judge is responsible for in that year. As alternatives to assess robustness, we look at number of sentences written and number of characters written.

Work quality is measured by the number of citations to a judge's opinions. Judges in a common-law system cite previous cases that are useful to their decision, and therefore citations can be seen as an expert evaluation of peer decision quality (Posner, 2008). More citations means that a case (and the authoring judge) have a stronger influence on the path of the law.

The citations measure is per case (divided by number of cases), so it is workload-adjusted. Citations are annotated as positive, negative, or distinguishing by the data provider, so for the baseline we look only at positive citations. As alternative measures, we use all cites (including negative and distinguishing), discussion cites (where the case was discussed at length by the citing court), and out-of-state cites (only citations in other jurisdictions). Because state supreme court precedents have no bindingness in other states, out-of-state citations serve as an especially strong signal of legal usefulness or influence (Choi et al., 2010).

To check for the importance of caseload changes, we report the number of opinions written as an outcome. To help assess the relative importance of output and quality, we also report a measure of *work impact* – the total number of positive citations to a judge in a year (unadjusted for number of opinions). The appendix includes a range of other outcomes, including measures of caselaw research and number of discretionary opinions written.

This approach faces the challenge that the difficulty and importance of a ruling varies from case to case, for reasons outside a judge's control. This problem is exacerbated when

trying to compare judges across different states (as done in Choi et al. 2010), since not only lower-court characteristics but also a range of court and state factors might affect number of opinions written, and citations to those opinions.

The goal of the analysis is to make two key comparisons: first, to compare a judge’s performance to other judges in the same court-year; and second, to compare a judge’s performance to his/her performance in other years. A challenging feature of the data is that the distributions of the outcomes are extremely variable across courts, judges, and years. This means that, when making within-court-year comparisons, for example, court-years with higher variance are upweighted in regressions using the raw data. In addition, coefficients on treatments that affect different subsets of states will not be comparable, as the different subsets will have different outcome variance.

We address this issue by normalizing outcome variables within subsets of observations. For the purposes of within-court-year comparisons (selection system and election cycle), we divide the outcomes for each judge-year by the court-year standard deviation, meaning that each court-year subsample will have variance one. In turn, for within-judge comparisons (retention system), outcomes are divided by the judge’s standard deviation, meaning that each judge’s sample of observations, together, will have variance one.⁹ We do not de-mean the outcomes within the groups; de-meaning does not change the coefficients or standard errors, but it does reduce adj. R^2 substantially. In the replication notebook we show that the state-year fixed effects and judge fixed effects each explain, by themselves, about 60% of the variance in both outcomes (output and quality). Together, they explain about 80% of the variance in output and 70% of the variance in quality.

Citations are a joint measure of work quality and case importance. Some types of cases are more important than others. For example, cases that review the constitutionality of statutes are probably relatively important. In addition, judges have some discretion over the types of cases they are chosen to author opinions for. If we want to compare the quality of judges working on the same court at the same time, we need to try to account for these non-judge factors.

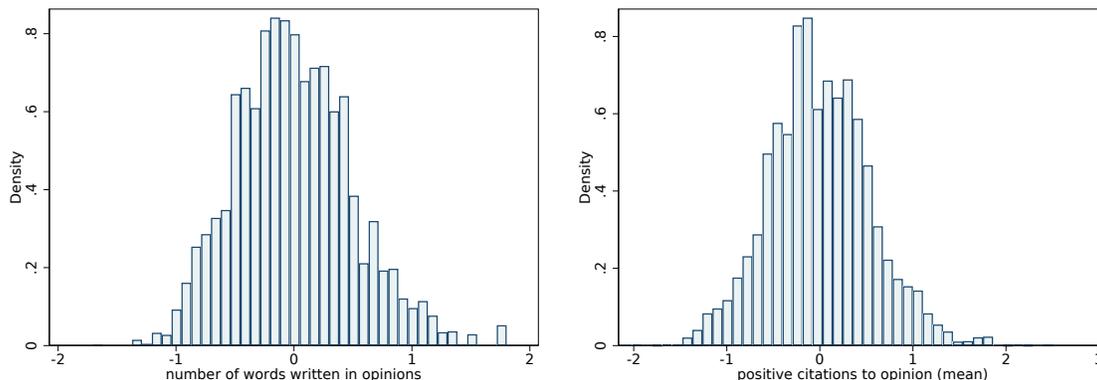
Empirically, we use the full range of dummy variables for the area of law of a case, as well as the related industries of a case. These are coded for each case by Bloomberg staff

⁹The main implication for our results is that we can compare the election-year effect sizes for partisan elections and nonpartisan elections. Without this normalization, the non-partisan effect appears larger than the partisan effect, but that turns out to be solely due to different variances in work output for these states. The only finding that is different with variance normalization is the effect of partisan elections on quality; without the normalization, this coefficient is still negative but not statistically significant. The full set of results without variance normalization are reported in the appendix.

Table 3: Summary Statistics on Area of Law and Related Industries

Area of Law	Freq.	Percent	Related Industrial Sector	Freq.	Percent
Criminal Law	191810	21.85	Real Estate	28527	13.64
Civil Procedure	74757	8.52	Law Enforcement	10758	5.14
Evidence	66377	7.56	Automobiles	10206	4.88
Torts	57915	6.6	Insurance	9158	4.38
Damages & Remedies	45073	5.14	Tax	8509	4.07
Contracts	40888	4.66	Construction & Engineering	6332	3.03
Real Property	36408	4.15	Workers' Compensation	5397	2.58
Constitutional Law	34038	3.88	Banking	4917	2.35
Family Law	32191	3.67	Legal & Compliance Services	4682	2.24
Workers' Compensation	22955	2.62	Automobile Insurance	4124	1.97
Insurance Law	19375	2.21	Property Management	4108	1.96
Administrative Law	18264	2.08	Transportation	3890	1.86
Wills, Trusts & Estates	18179	2.07	Child Welfare	3689	1.76
Tax & Accounting	16978	1.93	Employment Services	3679	1.76
Employment Law	14601	1.66	Health & Medical	3478	1.66
Habeas Corpus	13426	1.53	Oil & Gas	3189	1.52
Appellate Procedure	13140	1.5	Railroads	2777	1.33
Professional Responsibility	12052	1.37	Hospitals	2719	1.3
Motor Vehicles & Traffic Law	9644	1.1	Education	2586	1.24
Land Use Planning & Zoning	9122	1.04	Trucking	2097	1
Government	8942	1.02	Bridges & Roads	1751	0.84
Mortgages & Liens	7531	0.86	Agriculture & Farming	1729	0.83
Landlord & Tenant	5499	0.63	Mortgage Lending	1680	0.8
Construction Law	4997	0.57	Manufacturing	1612	0.77
Elections & Politics	4972	0.57	Real Estate Agents & Brokers	1573	0.75
Eminent Domain	4943	0.56	Unions	1485	0.71
Labor Law	4790	0.55	Financial Services	1469	0.7
Government Employees	4773	0.54	Judiciary	1448	0.69
Debtor Creditor	4260	0.49	Politics	1336	0.64
Employee Benefits	4208	0.48	Teachers	1300	0.62
Medical Malpractice	4113	0.47	Medical Procedures	1273	0.61
Personal Property	3994	0.46	Public Works	1223	0.58
Corporate Law	3958	0.45	Life Insurance & Annuities	1155	0.55
Negotiable Instruments	3843	0.44	Apartment Leasing	1127	0.54
Education Law	3803	0.43	Mining & Natural Resources	1115	0.53
Banking & Finance	3380	0.39	Drug Trafficking	1105	0.53
Alcohol & Beverage	3213	0.37	Sewer & Water	990	0.47
Civil Rights	3138	0.36	Electric	985	0.47
Health Law	2950	0.34	Water & Sewer	972	0.46
Transportation Law	2839	0.32	Physicians	966	0.46
Partnerships	2333	0.27	Firearms & Weapons	962	0.46
Natural Resources	2301	0.26	Motorcycles	919	0.44
Legal Malpractice	2285	0.26	Water	904	0.43
Products Liability	2280	0.26	Food & Beverage	888	0.42
Alternative Dispute Resolution	2144	0.24	Commercial Real Estate	883	0.42
Communications & Media	2048	0.23	Property & Casualty Insurance	854	0.41
Environmental Law	1857	0.21	Administration	837	0.4

Figure 1: Distribution of Judge Fixed Effects for Output and Quality



Notes. Judge-specific means for Work Output (number of words written in opinions, left panel) and Work Quality (positive citations per opinion, right panel), after residualizing on court-year fixed effects and case characteristics.

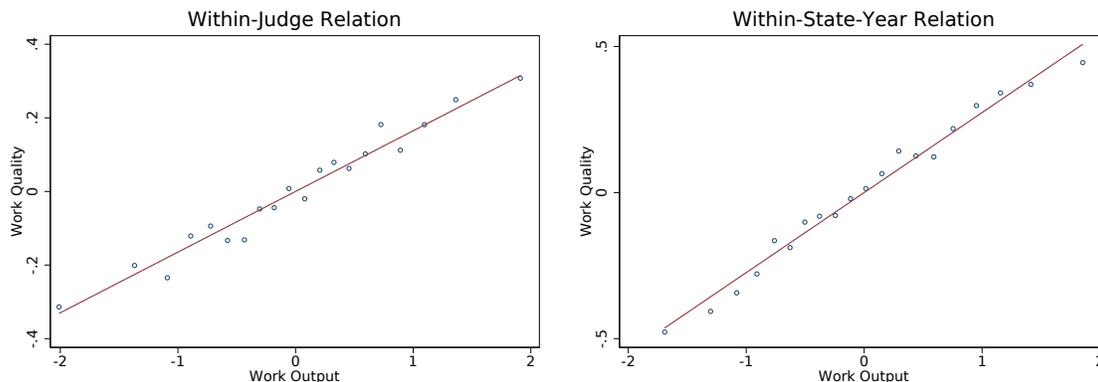
attorneys, and there may be up to three legal areas and three related industrial sectors for any particular case. Summary tabulations for the most frequent legal areas and industrial sectors are reported in table 3. The case characteristics vector includes a dummy variable for each area and sector, equaling one if the case has been assigned to that area or sector. Because there are so many of these characteristics, including separate covariates for every category would almost saturate the dataset. Instead, we include the first five principal components of this matrix of controls, which explains 65% of the variance of the matrix of case controls.¹⁰

An important issue is that a “judge” is not really a single individual, but a team of individuals that includes clerks and secretarial staff. Judges select the clerks that are working for them, and hence our measures can be seen as composites that depend upon both the judge’s legal skill when researching, reasoning, and writing, as well as managerial skill when selecting and directing clerks. As we know from Bloom et al. (2012), management quality varies across firms, and there are systematic relationships between management quality and firm performance.

In our data we cannot directly disentangle managerial skill from legal skill. However, we can ask if there is variation across judges in the same court, and/or across time within-judge. Figure 1 demonstrates the variation left over in work output and work quality after residual-

¹⁰Using more or fewer components does not change anything. See the appendix for details. In the appendix, we report results with log transformations of the outcomes. Taking logs does not change any of the main results. It tends to strengthen significance in most cases.

Figure 2: Output and Quality are Positively Correlated



Notes. Binscatters for Work Quality (positive citations per opinion, vertical axis) plotted conditional on Work Output (number of words written in opinions, horizontal axis). In left panel, values are standardized within judge. In right panel, values are standardized within state year.

izing out state-year fixed effects and case characteristics. This is the outcome variation left over for use in the estimation of treatment effects. As can be seen in the figure, there is significant variation in output and quality across judges, even after controlling for institutional and case-level characteristics. Next, Figure 2 shows that our performance measures are positively correlated, both within judge over time, and within court-year between judges. In the appendix we report plots of our judge quality measures over time for individual judges, to illustrate that the measures can reliably distinguish individuals over time.

To validate our outcome variables as judge-specific measures of performance, we explore the extent to which they are correlated with another performance measure previously used in the literature, the quality ratings issued by state bar associations. We were able to merge our data for a small number of judges with the data on evaluations provided by Lim and Snyder (2015b). We then regressed our performance measures on the bar evaluations with state-year fixed effects, to see whether our quality and output measures are predictive of the bar association evaluations of a “good judge,” as coded by the authors. Table 4 Column 3 shows that quality, but not output, is a strong predictor of the bar association qualifications evaluation.¹¹

¹¹An important caveat is that this difference is only identified off of three states where there were both “good” and “not-good” supreme court judges during the time overlap (1987 through 1994). A second caveat is the evidence in ? that attorney evaluations of judges seem to reflect partisan political affiliations and ideological differences, rather than judicial performance.

Table 4: Quality and Output as Predictors of Bar Association Evaluations

	(1)	(2)	(3)
	<i>Bar Evaluation: "Good Judge"</i>		
Work Output		1.313 (1.332)	-0.0332 (2.709)
Work Quality	2.633 (0.0483)		2.642 (0.820)
N	61	61	61
Court-Year FE's	X	X	X

Notes. $N = 61$ judge-years for these set of judges in Pennsylvania, Texas, and Washington for the years 1987 through 1994. Outcome is an indicator for being a "good" judge as defined in Lim and Snyder (2015), with mean 0.86. Coefficients estimated by conditional logit. Standard errors clustered by state in parentheses.

5 Econometric Approach and Results

The empirical analysis follows the theoretical analysis. In turn, we look at election-year pressures, then changes to the retention system, and then finally to the selection system.

5.1 Effect of Election-Cycle Pressure on Incumbent Judges

How do judges change their behavior over time in response to the election cycle? Ash and MacLeod (2015) show that contested elections reduce judicial performance. We add to that analysis by distinguishing between partisan and nonpartisan elections. In theory, if judges wish to be re-elected then they should put effort into election-year politics. This in turn would lead to a reduction in time spent on judging.

The empirical strategy for examining the effects of electoral demands on judicial behavior is to exploit the staggered election cycle for identification of stronger electoral incentives. The election schedule is arbitrarily assigned by history, so it is reasonable to assume that the schedule is uncorrelated with other institutional or socioeconomic factors that might affect individual judge performance. For this analysis we used data provided by Kritzer (2011), supplemented by new data collection and checking by a team of research assistants.

The electoral cycle is represented in our regressions as a vector of dummy variables E_{jst} , which equals one if judge j in state s is up for election at year t . The vector includes separate indicators for partisan, nonpartisan, and uncontested retention elections. The dummy

variable is coded as a one regardless of whether the judge actually ran for election – this is intended to avoid endogeneity problems from the judge’s choice whether to actually run. Similar results were obtained using an intensity treatment variable giving the number of years until the next election.

One possible source of bias in this analysis comes from time-invariant characteristics of individual judges. Some judges may have higher or lower performance than others on average due to unobservable characteristics, and they may be up for election more often or less often for any number of reasons. To deal with this possibility, we include a full set of judge-specific fixed effects. Therefore any estimated election coefficients are relative to a judge’s personal average.

A second major source of bias comes from the time-varying changes in the court work environment which may be correlated with the electoral schedule. For example, there may be campaigning demands during election years on all judges – not just those up for election – if they are asked to assist fellow members of their political party. To deal with this possibility, we include a full set of court-year fixed effects. Therefore any estimated election coefficients are also relative to the court average in each year. This means they effectively compare judges sitting on the same court, working at the same time, but who are in different stages of the electoral cycle.

Our preferred specification is

$$y_{jst} = \text{JUDGE}_j + \text{STATE}_s \times \text{YEAR}_t + E'_{jst}\rho + \epsilon_{jst} \tag{5.1}$$

where JUDGE_j is a judge fixed effect, $\text{STATE}_s \times \text{TIME}_t$ is a court-year fixed effect for each state s and year t , and E_{jst} includes the election-year treatments. Standard errors are clustered by state. Note that this gives the average output deviation for the year before an election.

The coefficient estimates from Equation (5.1) are reported in Table 5. In Columns 1 through 3, we see that both partisan elections and nonpartisan elections have a negative election-year effect on work output. This effect does not change much when adding judge fixed effects, case topic controls, or judge experience controls. In all cases, there is no negative effect of uncontested retention elections on work output; if anything there is a positive effect.

We don’t see much of an effect of election-year pressure on work quality (Columns 4 through 6). In nonpartisan and uncontested elections, the effects are clearly zero. In partisan elections, there is a negative coefficient but it is not significant.

The effects on work output observed under partisan and nonpartisan elections are driven

Table 5: Effect of Election-Year Pressure on Judge Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<i>Work Output</i>			<i>Work Quality</i>				
							<i># Cases</i>	<i>Impact</i>
Partisan	-0.142	-0.159	-0.116	0.0357	-0.0498	-0.0481	-0.0721	-0.110
Election Year	(0.067)	(0.051)	(0.042)	(0.048)	(0.037)	(0.035)	(0.029)	(0.042)
	[0.040]	[0.0033]	[0.0078]	[0.46]	[0.19]	[0.17]	[0.016]	[0.013]
Nonpartisan	-0.151	-0.165	-0.0906	-0.0122	-0.0550	-0.0537	-0.144	-0.105
Election Year	(0.067)	(0.068)	(0.036)	(0.065)	(0.065)	(0.066)	(0.070)	(0.031)
	[0.029]	[0.020]	[0.015]	[0.85]	[0.40]	[0.42]	[0.045]	[0.0015]
Uncontested	0.0563	0.0498	0.0314	0.0425	0.0137	0.00802	-0.0191	0.00205
Election Year	(0.061)	(0.062)	(0.046)	(0.074)	(0.062)	(0.061)	(0.055)	(0.068)
	[0.36]	[0.42]	[0.50]	[0.57]	[0.83]	[0.90]	[0.73]	[0.98]
N	14890	14890	14890	14890	14890	14890	14885	14890
Adj. R^2	0.723	0.832	0.880	0.972	0.982	0.982	0.956	0.923
Court-Year FE's	x	x	x	x	x	x	x	x
Judge FE's		x	x		x	x	x	x
Topic Controls			x			x	x	x
Exp. Controls			x			x	x	x

Treatment variable is a dummy equaling one for the years where a judge is facing reelection. Topic controls include area of law and related industries. Experience controls include a full range of dummies for half-decades of experience. See outcome descriptions in text. Outcomes normalized to have variance one within court-year. Standard errors, adjusted for two-way clustering by state and year, in parentheses. P-values in brackets.

Table 6: Effect of Election-Year Pressure, Additional Performance Measures

	(1)	(2)	(3)	(4)	(5)
	<i>Work Output</i>		<i>Work Quality (Citations)</i>		
	<i>Characters</i>	<i>Sentences</i>	<i>All</i>	<i>Discuss</i>	<i>Out-State</i>
Partisan Election Year	-0.114 (0.041) [0.0079]	-0.132 (0.041) [0.0021]	-0.0529 (0.030) [0.087]	-0.0404 (0.042) [0.34]	-0.0177 (0.045) [0.70]
Nonpartisan Election Year	-0.0917 (0.031) [0.0051]	-0.0872 (0.039) [0.032]	-0.0683 (0.055) [0.22]	-0.123 (0.062) [0.052]	-0.0218 (0.061) [0.72]
Uncontested Election Year	0.0315 (0.050) [0.53]	0.0196 (0.045) [0.66]	0.0116 (0.063) [0.86]	0.0611 (0.049) [0.22]	0.0982 (0.077) [0.21]
N	14890	14890	14890	14890	14886
Adj. R^2	0.912	0.885	0.964	0.911	0.980
Court-Year FE's	x	x	x	x	x
Judge FE's	x	x	x	x	x
Topic Controls	x	x	x	x	x
Exp. Controls	x	x	x	x	x

Treatment variable is a dummy equaling one for the years where a judge is facing reelection. Topic controls include area of law and related industries. Experience controls include a full range of dummies for half-decades of experience. See outcome descriptions in text. Outcomes normalized to have variance one within court-year. Standard errors, adjusted for two-way clustering by state and year, in parentheses. P-values in brackets.

at least in part by a decrease in workload. Column 7 shows that they are ruling on significantly fewer cases than their colleagues when up for election. Due to the decrease in cases published, the election-cycle judges have significantly lower impact in terms of total citations per year (Column 8).

Table 6 extends the analysis to the alternative specifications for performance. These results are in line with what we saw in the previous table. First, from Columns 1 and 2 we learn that the electoral effect on output is not sensitive to how output is defined. Second, we see negative coefficients for the effect of partisan elections on quality, although only “all” citations is significant.

A visualization of the results and a range of robustness checks and alternative specifi-

cations are reported in the appendix. We got similar results after dropping short opinions (less than 3 paragraphs). We also got similar results when looking at different types of cases (criminal, civil, administrative, or constitutional), when using logs of the outcomes, including more case factor controls, or dropping the first and last year for each judge career. Without normalizing the variance by state-year, or when normalizing by judge, the effects of partisan elections on quality goes to zero. When weighting by number of opinions, the effect of non-partisan elections is larger, and the effect of partisan elections is smaller and not significant.

The election-cycle results suggest that contested elections reduce output, while uncontested elections have no effect. Since uncontested elections are non-competitive by design, the null effect here is unsurprising.¹² In addition, partisan elections tend to reduce work quality, but not nonpartisan elections. These results are consistent with the hypothesis that partisan judges have a less preference for producing high quality decisions. The fact that in election years there is a fall in performance is consistent with the hypothesis that voters do care, and that judges do need to spend time to get re-elected.

5.2 Effect of Retention Process on Incumbent Judges

This subsection reports the results on how changing the system for judge retention affects the performance of sitting judges. The theory predicts that judges will work to get re-elected, though their choice between pandering and providing information regarding their work quality depends upon which factor is more important for their re-election chances. We identify the effect of the re-election system using discrete changes in the rules for retaining state supreme court judges. Four states moved from partisan retention elections to nonpartisan retention elections: Florida, Georgia, Kentucky, and Utah. Eight states moved from partisan retention to uncontested retention elections: Colorado, Illinois, Iowa, Indiana, Kansas, Nebraska, New Mexico, and Oklahoma. Six states moved from nonpartisan retention to uncontested retention: Arizona, Florida, Maryland, South Dakota, Utah, and Wyoming.¹³

The regression framework is a standard differences-in-differences approach based on Bertrand et al. (2004). To control for time-invariant judge characteristics that may be correlated with the retention system in various states, we include judge fixed effects. To control for national trends in performance, we include year fixed effects. To control for pre-

¹²Note that since the mid-1990s, third-party funding for negative advertising in Missouri Plan elections has increased significantly. Our results may not extend to more recent years (our panel ends in 1994). This is an important area for future research.

¹³See notes in Subsection 5.1, and see Appendix B.1 for more details and discussion of robustness checks.

existing state trends in performance that may be confounded with the reforms, we include state-specific linear trends. In the strictest specification, we include controls for a number of other rules affecting the judiciary, notably state expenditures on courts.

We measure effects in a ten-year window before and after the reforms. The regressions include an indicator equaling one for the baseline time window of ten years before and ten years after a change to the retention system. The treatment variable is a dummy for the ten years after the change. Thus, with the inclusion of the judge fixed effects, the estimates can be interpreted as the average difference in within-judge performance for the ten years after the policy change relative to the ten years before the policy change. In a handful of states, we shrank the time window if the reform occurred close to the beginning or end of the sample.¹⁴ A graphical representation of the windows is included in Appendix Figure 3. In the appendix we include a table using other time windows; our main result on nonpartisan judges and quality is a somewhat lagged effect that is statistically significant with an effect window of at least eight years.

Formally, we estimate

$$y_{jst} = \text{YEAR}_t + \text{JUDGE}_j + \text{STATE}_s \times t + \bar{R}'_{st}\bar{\rho} + R'_{st}\rho + X'_{jst}\beta + \epsilon_{jst} \quad (5.2)$$

where YEAR_t is a fixed effect for the two-year period t , JUDGE_j is a judge fixed effect, and $\text{STATE}_s \times t$ is a state-level linear time trend for state s . The term \bar{R}_{st} is a vector of indicators equaling one for the baseline time windows of ten years before and ten years after each of the retention reforms. R_{st} is a vector of treatment indicators for the ten years after each rule change (with ρ measuring the corresponding causal effects of interest). X_{jst} includes other state and judge controls, namely variables for case topic and other court-related policies. Standard errors are clustered by state and year.

Table 7 reports the estimates for ρ from Equation 5.2. First, in Columns 1 through 3 we look at effects of the rule changes on average judge output. There is an inconsistently significant negative effect of moving from partisan to nonpartisan elections. Effects for partisan-to-uncontested and nonpartisan-to-uncontested, change a lot across specifications.

Next, Columns 4 through 6 show no effect of moving from partisan to nonpartisan elections, or moving from partisan to uncontested, on work quality. However, there is a positive

¹⁴These reforms are mostly enacted by voters through ballot referendums administered in November and officially going into effect the subsequent January. In these cases the dummy variable would turn on in the year following the vote. In cases where the policy is effective in the first half of the year, it is coded as turning on in that year. Note that Florida changes from partisan to nonpartisan and then to uncontested elections. Our results do not change substantially if Florida is left out of the analysis.

Table 7: Effect of Judicial Retention System on Judge Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<i>Work Output</i>			<i>Work Quality</i>			<i># Cases</i>	<i>Impact</i>
Partisan to	-0.502	-0.562	-0.153	0.179	0.178	0.104	0.0160	0.0285
Nonpartisan	(0.27)	(0.26)	(0.15)	(0.26)	(0.28)	(0.23)	(0.28)	(0.25)
	[0.069]	[0.033]	[0.30]	[0.49]	[0.52]	[0.66]	[0.96]	[0.91]
Partisan to	-0.0117	0.0417	0.309	0.0531	0.0190	-0.0829	0.306	0.171
Uncontested	(0.085)	(0.11)	(0.059)	(0.13)	(0.13)	(0.16)	(0.098)	(0.10)
	[0.89]	[0.70]	[0.0001]	[0.69]	[0.89]	[0.60]	[0.003]	[0.094]
Nonpartisan to	-0.151	-0.217	0.0595	0.389	0.378	0.369	-0.0571	0.118
Uncontested	(0.16)	(0.15)	(0.15)	(0.18)	(0.19)	(0.16)	(0.11)	(0.16)
	[0.36]	[0.16]	[0.68]	[0.037]	[0.052]	[0.022]	[0.61]	[0.46]
N	13756	13756	13756	13756	13756	13756	13746	13754
Adj. R^2	0.855	0.860	0.909	0.987	0.987	0.987	0.902	0.946
Judge + Year FE's	x	x	x	x	x	x	x	x
State Trends		x	x		x	x	x	x
Topic Controls			x			x	x	x
Rule controls			x			x	x	x

Topic controls include area of law and related industries. State policy controls includes other appointment-process changes, mandatory-retirement changes, changes in the number of judges, and log state government expenditures on the judicial branch. Outcomes normalized to have variance one within judge. Standard errors, adjusted for two-way clustering by state and year, in parentheses. P-values in brackets.

Table 8: Effect of Judicial Retention System, Additional Performance Measures

	(1)	(2)	(3)	(4)	(5)
	<i>Work Output</i>		<i>Work Quality (Citations)</i>		
	<i>Characters</i>	<i>Sentences</i>	<i>All</i>	<i>Discuss</i>	<i>Out-State</i>
Partisan to Nonpartisan	-0.561 (0.26) [0.034]	-0.541 (0.28) [0.058]	0.167 (0.27) [0.54]	0.0151 (0.20) [0.94]	-0.0194 (0.17) [0.91]
Partisan to Uncontested	0.0493 (0.16) [0.76]	0.0298 (0.17) [0.86]	0.0539 (0.13) [0.69]	-0.0331 (0.11) [0.76]	0.0615 (0.11) [0.59]
Nonpartisan to Uncontested	-0.219 (0.32) [0.50]	-0.203 (0.40) [0.61]	0.436 (0.20) [0.037]	0.219 (0.24) [0.37]	0.211 (0.18) [0.25]
N	13756	13756	13756	13752	13754
Adj. R^2	0.844	0.883	0.901	0.888	0.927
Judge + Year FE's	x	x	x	x	x
State Trends	x	x	x	x	x
Topic Controls	x	x	x	x	x
Rule controls	x	x	x	x	x

Topic controls include area of law and related industries. State policy controls includes other appointment-process changes, mandatory-retirement changes, changes in the number of judges, and log state government expenditures on the judicial branch. Outcomes normalized to have variance one within judge. Standard errors, adjusted for two-way clustering by state and year, in parentheses. P-values in brackets.

and statistically significant effect when moving from a nonpartisan system to uncontested elections. This result is insensitive to the addition of state time trends, judge experience controls, and a range of state-level policy controls.

In terms of caseload (Column 7) and total citations (Column 8), there are no effects for partisan to nonpartisan, nor for nonpartisan to uncontested. We see positive effects for partisan to uncontested, but these are sensitive to the inclusion of trends and controls.

Finally, Table 8 adds to the results with the alternative outcomes. In terms of output, we see more consistent negative effects for moving from partisan to nonpartisan elections. In terms of quality, there is a significant positive effect on “all” citations for the nonpartisan-to-uncontested reform (Column 3). For discussion cites (Column 4) and out-of-state cites (Column 5), the coefficients are positive but not significant at the standard thresholds.

In the appendix and replication materials we provide other robustness checks and alternative specifications. We got similar results after dropping short opinions (less than 3 paragraphs). The quality results were similar in magnitude, but more significant, when weighting by number of opinions written, using log transformations of the outcome, or adding more case factor controls. Dropping the first and last year for each judge reduced significance of the main quality regression. The retention effect is statistically significant with an effect window of at least 8 years.¹⁵ There was no effect on probability of being overruled by the U.S. Supreme Court. Finally, we dropped each treated state one by one; the nonpartisan-to-uncontested effects on quality were significant in all regressions.

These results provide additional evidence that electoral incentives have an important impact on the performance of appellate court judges. In particular, the partisanship of elections matters. When moving from partisan to nonpartisan elections, there is a reduction in output, consistent with the hypothesis that nonpartisan elections are more competitive. When taking nonpartisan judges and giving them tenure, meanwhile, there is an increase in work quality. This is consistent with the idea from Ash and MacLeod (2015) that these judges have an intrinsic motivation to produce high-quality work, and that with weaker electoral incentives they will orient their time and decision choices away from campaigning and pleasing voters, and toward writing good decisions that please their peers.

On the other hand, when taking partisan judges and giving them tenure, there is no effect on output or quality. This suggests, consistent with Judge O'Connor's view, that partisan elections are less competitive than nonpartisan elections. Because there was little electoral pressure in the first place, eliminating those elections does not have a big impact on the way partisan judges spend their time or decide their cases.

Second, these results are consistent with the hypothesis that partisan-selected judges have a lower intrinsic valuing of quality than non-partisan judges. When giving them tenure, they do not value quality enough to invest in it with the newly available time. Because we found above that partisan election cycles reduce output (and therefore seem to involve election pressure), the evidence is more in line with this selection mechanism.

¹⁵This delayed effect is likely due to the relatively long term lengths for supreme court judges. The minimum term length for these treated states is six years, the maximum is ten years, and the average is 7.8 years. Therefore one would expect a delay in the incentive effect until the next election, which could be as long as ten years.

5.3 Effect of Selection Process on Selected Judges

The evidence on election year effort shows that election pressure does count, and that potential judges do attempt to influence the vote. The evidence is consistent with the hypothesis that judges selected under a non-partisan systems place more value upon quality than quantity relative to a judge selected under a non-partisan system. This weighting does not necessarily imply that they are worse judges. The effect of the selection system upon overall quality of a judge can be measured in our data by measuring the effect of a change in a state appointment system upon judges selected under different systems but working on the same court.

In Table 1 we see that four states changed from partisan selection to nonpartisan selection: Florida, Georgia, Kentucky, and Utah.¹⁶ Seven states moved from partisan selection to merit selection: Colorado, Iowa, Indiana, Kansas, Nebraska, and Oklahoma.¹⁷ Six states moved from nonpartisan selection to merit selection: Arizona, Florida, Maryland, South Dakota, Utah, and Wyoming.¹⁸ The goal is to compare the performance of judges selected before these reforms to the performance of judges selected after these reforms.

We control for time-varying state-specific factors by including a full set of court-year (interacted) fixed effects. This specification effectively compares the performance of judges sitting on the same court at the same time, but selected under different regimes. We include case characteristics to address the issue of endogenous selection of case types to judge types. In the strictest specification, we include a set of covariates for judge personal characteristics. These include controls for judge starting age, gender, political party, and whether they come from a top law school.

The estimating equation for performance variable y_{jst} for judge j in state s at year t is

$$y_{jst} = \text{STATE}_s \times \text{YEAR}_t + X'_{jst}\beta + S'_{jst}\rho + \epsilon_{jst} \quad (5.3)$$

where $\text{STATE}_s \times \text{YEAR}_t$ includes the court-year fixed effects, X_{jst} includes the case and judge controls, and S_{jst} includes the treatment indicators equaling one for judges selected under the post-reform system.¹⁹ Given the inclusion of the fixed effects, the coefficients ρ procure

¹⁶Florida moved from nonpartisan to merit five years after the partisan to nonpartisan reform, and only four judges were selected under the nonpartisan system. Therefore Florida is excluded from the baseline selection regressions, but including it does not change the results.

¹⁷Tennessee moved to merit selection in 1972, but moved back to partisan selection in 1978. Only one judge was selected by the merit process so it is not included in the analysis.

¹⁸See Footnote 16 re Florida.

¹⁹Note that in the electoral selection systems, the judges may be initially appointed by the governor to

the average difference in performance between judges selected under the new system and judges selected under the old system, controlling for other time-varying state-level factors. Standard errors are clustered by state and year, where generally we got smaller standard errors with clustering by state, clustering by state-year, or not clustering.

Table 9 reports the estimates from Equation (5.3) for our baseline set of outcomes. First, in Columns 1 through 3, we do not see any robust selection effect on output. Judges selected by partisan elections, non-partisan elections, and merit commissions write about the same number of words per year.

In Columns 4 through 6, however, we see significant effects on work quality. Relative to their partisan-elected colleagues, nonpartisan-elected judges write higher-quality decisions. In turn, merit-selected judges also write higher-quality decisions than partisan-elected colleagues. This effect is not driven by differences in the types of cases they are assigned (Column 5). It is also not explained by judge demographic characteristics (Column 6).

As seen in Column 7, there is no significant difference between judges in terms of the number of opinions they are assigned. As a result, the effect on total citations to a judge in a year (Column 8) is positive for both nonpartisan relative to partisan, and merit relative to partisan. Meanwhile, there is a marginally significant negative difference between merit-selected judges and nonpartisan-elected colleagues.

Table 10 reports the selection-system effects for some additional measures of output and quality. In Columns 1 and 2, we see again that there is no difference between the types of judges in terms of output. Columns 3 through 5 show that in the case of quality, there are robust differences between the systems. When looking at all citations (including negative and distinguishing, not just positive) or discussion citations (where the previous cases is actively discussed), there are significant positive differences between nonpartisan and partisan, and between merit and partisan. In the case of out-of-state citations, the effect of nonpartisan (from partisan) is no longer significant, but strongly significant for merit selection (relative to partisan).

We ran a number of robustness checks and alternative specifications, which are reported in the replication materials. In general, the results on partisan-to-merit are more robust than the results on partisan-to-nonpartisan. The results were not sensitive to subsetting by type of case (criminal, civil, etc.), using log transformations, dropping first and last year of each judge career, or adding more case characteristics components. Without normalizing

fill a vacant seat, rather than being initially selected through a competitive electoral process. We still code the appointed judges as being selected under the electoral system – since the predecessor’s choice whether to step down is endogenous to the system.

Table 9: Effect of Judicial Selection System on Judge Performance

	<u>Work Output</u>			<u>Work Quality</u>			<u># Cases</u>		<u>Impact</u>
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	
Nonpartisan, Relative to Partisan	0.165 (0.027) [0.0001]	0.0306 (0.070) [0.66]	-0.000794 (0.075) [0.99]	0.326 (0.089) [0.0006]	0.300 (0.091) [0.0020]	0.225 (0.11) [0.038]	0.186 (0.15) [0.21]	0.250 (0.13) [0.054]	
Merit, Relative to Partisan	0.198 (0.11) [0.083]	0.194 (0.11) [0.084]	0.193 (0.10) [0.06]	0.457 (0.16) [0.0051]	0.452 (0.15) [0.0047]	0.512 (0.14) [0.0006]	0.0307 (0.073) [0.67]	0.332 (0.064) [0.0001]	
Merit, Relative to Nonpartisan	-0.0160 (0.15) [0.92]	-0.0089 (0.13) [0.95]	0.00399 (0.14) [0.98]	0.00177 (0.24) [0.99]	0.0249 (0.24) [0.92]	0.0506 (0.22) [0.82]	0.00622 (0.069) [0.93]	-0.0431 (0.15) [0.78]	
Start Age		0.00229 (0.0033)				-0.0109 (0.0027)			
Female		0.00879 (0.056)				0.0694 (0.077)			
Democrat		-0.0402 (0.045)				-0.0455 (0.046)			
Republican		0.0726 (0.058)				-0.0255 (0.052)			
Top School		0.0337 (0.058)				0.0114 (0.058)			
N	14996	14996	14996	14996	14996	14996	14991	14996	
Adj. R^2	0.723	0.806	0.806	0.972	0.972	0.973	0.948	0.894	
Court-Year FE's	x	x	x	x	x	x	x	x	
Topic Controls		x	x	x	x	x	x	x	

Estimates of the average difference between judges selected under a new system, relative to judges selected under the old system. Outcomes normalized to have variance one within court-year. Standard errors, adjusted for two-way clustering by state and year, in parentheses. P-values in brackets.

Table 10: Effect of Judicial Selection System, Additional Performance Measures

	(1)	(2)	(3)	(4)	(5)
	<i>Work Output</i>		<i>Work Quality (Citations)</i>		
	<i>Characters</i>	<i>Sentences</i>	<i>All</i>	<i>Discuss</i>	<i>Out-State</i>
Nonpartisan, Relative to Partisan	0.0251 (0.070) [0.72]	0.0833 (0.17) [0.62]	0.303 (0.084) [0.00073]	0.340 (0.089) [0.0004]	0.192 (0.16) [0.25]
Merit, Relative to Partisan	0.192 (0.11) [0.088]	0.206 (0.11) [0.069]	0.453 (0.15) [0.0042]	0.372 (0.11) [0.0014]	0.353 (0.082) [0.0001]
Merit, Relative to Nonpartisan	-0.00370 (0.13) [0.98]	0.00195 (0.11) [0.99]	0.0334 (0.25) [0.89]	0.0757 (0.19) [0.70]	-0.0714 (0.12) [0.55]
N	14996	14996	14996	14996	14992
Adj. R^2	0.858	0.819	0.946	0.880	0.976
Court-Year FE's	x	x	x	x	x
Topic Controls	x	x	x	x	x

Estimates of the average difference between judges selected under a new system, relative to to judges selected under the old system. Outcomes normalized to have variance one within court-year. Standard errors, adjusted for two-way clustering by state and year, in parentheses. P-values in brackets.

the variance within court year, some of the results on work quality are only significant at the 10%, rather than 1% or 5%, level. The results were of similar magnitude, but more statistically significant, when weighting by number of opinions written. The results on partisan-to-nonpartisan were somewhat sensitive to the “window” before and after the form for including judges: if you only look at judges selected 5 years either side of the reform, the effect is not significant, but the effect is seen when including judges selected at least 10 years either side of the reform. Adding state-specific trends in judge starting year shrinks the coefficients; the effects are significant only with clustering by state-year (rather than two-way clustering by state and year). When requiring that there be at least 2 judges selected from each system, the effect of partisan-to-nonpartisan shrinks and is no longer significant. Finally, we dropped each treated state one by one; the partisan-to-nonpartisan results were unaffected, while the partisan-to-merit effects shrank and became marginally insignificant when either Indiana or Oklahoma were dropped.

Overall, these results point to the importance of selection systems for the quality of judge decision-making. Partisanship of election processes reduces decision quality, but not total output. From A. Kirkland and Coppock (2017) we know that partisan labels leads to less scrutiny of the candidates. Hence this finding is consistent with the hypothesis that the primary system does not select better candidates than an at large election. In these regressions there seems to be little difference in quality between a non-partisan systems and merit systems.

6 Conclusion

The goal of this paper has been to contribute some evidence regarding the hypothesis that the choice between a “politician” and “bureaucrat” entails a trade-off between a sensitivity to the desires of the electorate and the execution of the mission to make high quality legal decisions (Maskin and Tirole, 2004; Alesina and Tabellini, 2007, 2008). A substantial body of work documents that public officials respond to the preferences of the electorate (Ashworth, 2012). We complement this literature with evidence on the other side of the balance sheet: how electoral pressures interact with a judge’s mission to provide high work quality. Sitting judges spend more time on their work in response to a weakening of re-election pressure, and they make decisions that are evaluated as higher-quality by other judges. Moreover, judges selected by a technocratic merit commission are of higher quality. The results are detailed in Table 11.

Table 11: Summary of Results

	<u><i>Effect on Output</i></u>	<u><i>Effect on Quality</i></u>
<u>Incentive Effects: Electoral Cycle</u>		
Partisan Election Year	Output ↓**	Quality ↓+
Nonpartisan Election Year	Output ↓**	~
Uncontested Election Year	~	~
<u>Incentive Effects: Retention Reform</u>		
Partisan to Nonpartisan	Output ↓+	~
Partisan to Uncontested	Output ↑+	~
Nonpartisan to Uncontested	~	Quality ↑*
<u>Selection Effects</u>		
Nonpartisan, Relative to Partisan	~	Quality ↑*
Merit, Relative to Partisan	~	Quality ↑**
Merit, Relative to Nonpartisan	~	~

Left-most column indicates the treatment, and the other column headers indicate the outcome measure. Arrows indicate a positive or negative effect on judge performance. A tilde (~) indicates no effect. Pluses (+) and asterisks (*, **) indicate significance/robustness of the effect.

For incumbent judges, we find that contested systems reduce output in election years, but not uncontested elections. This is consistent with a simple model in which campaign effort takes time away from judging. Moving from nonpartisan to uncontested elections increases case quality, consistent with the notion that nonpartisan contested elections are more demanding of a judge’s time than uncontested elections. There is no within-judge effect of moving from partisan to uncontested elections, reflecting that nonpartisan elections are most competitive – due to less bias than partisan elections. Finally, the merit-based selection process selects better judges than the election systems. These results are consistent with a selection model where better-informed experts can choose higher-quality officials than voters on average.

Our evidence is broadly in line with the early rational-choice approaches of Downs (1957) and Ferejohn (1986), in which voters use their information to make the best decisions they can, conditional upon their policy preferences. But more information is not always better; more information on candidate quality can improve performance (see Pande, 2011), but more information on political affiliation can reduce performance.

Should all states immediately move to a merit system with uncontested retention elections? Our evidence would certainly strengthen arguments to do so. But there are other criteria besides judicial citations for ranking courts, and ballot referenda for the merit plan have failed many times. There may be many other social impacts of these courts, but at present there aren’t data-driven ways to measure them. There is an ongoing debate on which is the superior system (e.g. Pozen, 2010); the fact that states continue to experiment with different systems suggests that it is not clear which system is optimal. If a single system were clearly optimal, then we would have expected the market to have moved in that direction quickly, consistent with Posner’s (1987) view that legal institutions move in the direction of efficient exchange.

The fact that we do find a pattern of effects predicted by our simple model helps explain why there is experimentation. While the results are consistent with merit commissions selecting better judges, judging is not a purely technical activity. The political views of judges color the ideological content of their decisions (see Epstein et al., 2013), which may explain why many jurisdictions prefer to give voters a clear signal of the political views of judges. Optimizing states would change systems only if it led to an improvement; hence at any point in time there should be only small variation across states (as Choi et al. (2010) find).²⁰

²⁰The issue here is similar to the well-known Roy model in economics. One cannot use cross-sectional data

Finally, our results highlight the fact that the American legal system is neither simple nor static. It is a complex, dynamic system consisting of a number of interlocking ingredients. Our study focuses upon one of the most important and influential ingredients of this system: the U.S. state supreme court judges who rule on all aspects of private law, including contract, tort, and property law. Our evidence is consistent with the hypothesis that these judges are professionals who are interested in enhancing the quality of the law. Hence, we have observed many states moving away from partisan political processes for selection toward nonpartisan and merit-based processes. These more “bureaucratic” systems have selected better judges and imposed incentives more aligned with the mission of increasing the quality of American law.

to measure the impact of a location on wages due to the self-selection by workers; similarly, one cannot use cross-sectional data to measure the impact of selection systems on judge performance due to self-selection of judges across states. See Heckman and Honore (1990) for details.

References

- A. Kirkland, P. and Coppock, A. (2017). Candidate choice without party labels:: New insights from conjoint survey experiments. *Political Behavior*, 40:1–21.
- Alesina, A. and Tabellini, G. (2007). Bureaucrats or politicians? part i: A single policy task. *The American Economic Review*, 97(1):169–179.
- Alesina, A. and Tabellini, G. (2008). Bureaucrats or politicians? part ii: Multiple policy tasks. *Journal of Public Economics*, 92(3-4):426–447.
- Alt, J., de Mesquita, E., and Rose, S. (2011). Disentangling accountability and competence in elections: Evidence from us term limits. *Journal of Politics*, 73(1):171–186.
- Ash, E. and MacLeod, W. B. (2015). Intrinsic motivation in public service: Theory and evidence from state supreme courts. *Journal of Law and Economics*, 58(4).
- Ashworth, S. (2012). Electoral accountability: Recent theoretical and empirical work. *Annual Review of Political Science*, 15(1):183–201.
- Ashworth, S., Bueno de Mesquita, E., and Friedenber, A. (2017). Accountability and information in elections. *American Economic Journal: Microeconomics*, 9(2):95–138.
- Ashworth, S. and de Mesquita, E. B. (2008). Electoral selection, strategic challenger entry, and the incumbency advantage. *Journal of Politics*, 70(4):1006–1025.
- Berdejo, C. and Yuchtman, N. (2013). Crime, punishment, and politics: an analysis of political cycles in criminal sentencing. *Review of Economics and Statistics*, 95(3):741–756.
- Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119(1):249–275.
- Besley, T. (2006). *Principled agents?: The political economy of good government*. Oxford University Press on Demand.
- Besley, T. and Case, A. (1995). Does electoral accountability affect economic-policy choices - evidence from gubernatorial term limits. *Quarterly Journal of Economics*, 110(3):769–798.
- Besley, T. and Coate, S. (2003). Elected versus appointed regulators: Theory and evidence. *Journal of the European Economic Association*, 1(5):1176–1206.
- Besley, T. and Payne, A. (2013). Implementation of anti-discrimination policy: Does judicial selection matter? *American Law and Economics Review*, 15(1):212–251.
- Bloom, N., Genakos, C., Sadun, R., and Reenen, J. V. (2012). Management practices across firms and countries. NBER Working Papers 17850, National Bureau of Economic Research, Inc.
- Bloom, N. and Reenen, J. V. (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics*, 122(4):1351–1408.

- Bonneau, C. W. and Cann, D. M. (2015). Party identification and vote choice in partisan and nonpartisan elections. *Political Behavior*, 37(1):43–66.
- Canes-Wrone, B., Clark, T. S., and Kelly, J. P. (2014). Judicial selection and death penalty decisions. *American Political Science Review*, 108:23–39.
- Canes-Wrone, B., Clark, T. S., and Park, J.-K. (2010). Judicial independence and retention elections. *Journal of Law, Economics & Organization*, 28(2):211.
- Choi, S. J., Gulati, G. M., and Posner, E. A. (2010). Professionals or politicians: The uncertain empirical case for an elected rather than appointed judiciary. *Journal of Law, Economics, and Organization*, 26(2):290.
- Condorcet, M. d. (1785). Essay on the application of analysis to the probability of majority decisions.
- Dewatripont, M., Jewitt, I., and Tirole, J. (1999). The economics of career concerns, part ii: Application to missions and accountability of government agencies. *Review of Economic Studies*, 66(1):pp.199–217.
- Downs, A. (1957). An economic theory of political action in a democracy. *Journal of Political Economy*, 65(2):pp. 135–150.
- Epstein, L., Landes, W. M., and Posner, R. A. (2013). *The Behavior of Federal Judges*. Harvard University Press.
- Fearon, J. D. (1999). Electoral accountability and the control of politicians: selecting good types versus sanctioning poor performance. *Democracy, accountability, and representation*, 55:61.
- Ferejohn, J. (1986). Incumbent performance and electoral control. *Public Choice*, 50(1-3):5–25.
- Ferraz, C. and Finan, F. (2008). Exposing corrupt politicians: The effects of brazil’s publicly released audits on electoral outcomes. *Quarterly Journal of Economics*, 123(2):703–745.
- Ferraz, C. and Finan, F. (2011). Electoral accountability and corruption: Evidence from the audits of local governments. *American Economic Review*, 101(4):1274–1311.
- Fournaies, A. and Hall, Andrew, B. (2018). How do electoral incentives affect legislator behavior? Technical report, Working Paper.
- Francois, P. (2000). Public service motivation as an argument for government provision. *Journal of Public Economics*, 78(3):275–299.
- Gordon, S. and Huber, G. (2007). The effect of electoral competitiveness on incumbent behavior. *Quarterly Journal of Political Science*, 2(2):107–138.
- Hall, M. (2007). Voting in state supreme court elections: Competition and context as

- democratic incentives. *Journal of Politics*, 69(4):1147–1159.
- Hall, M. and Bonneau, C. (2006). Does quality matter? challengers in state supreme court elections. *American Journal of Political Science*, 50(1):20–33.
- Hanssen, F. (2004). Is there a politically optimal level of judicial independence? *The American Economic Review*, 94(3):712–729.
- Hanssen, F. A. (1999). The effect of judicial institutions on uncertainty and the rate of litigation: The election versus appointment of state judges. *The Journal of Legal Studies*, 28(1):pp.205–232.
- Heckman, J. J. and Honore, B. E. (1990). The empirical content of the roy model. *Econometrica*, 58(5):pp.1121–1149.
- Helland, E. and Tabarrok, A. (2002). The effect of electoral institutions on tort awards. *American Law and Economics Review*, 4(2):341–370.
- Hirano, S. and Snyder Jr, J. M. (2014). Primary elections and the quality of elected officials. *Quarterly Journal of Political Science*, 9(4):473–500.
- Holmström, B. and Milgrom, P. (1991). Multi-task principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization*, 7:24–52.
- Huber, G. A. and Gordon, S. C. (2004). Accountability and coercion: Is justice blind when it runs for office? *American Journal of Political Science*, 48(2):247–263. Times Cited: 79
Huber, GA Gordon, SC Huber, Gregory/A-5950-2012 Huber, Gregory/0000-0001-6804-8148 79.
- Iaryczower, M., Lewis, G., and Shum, M. (2013). To elect or to appoint? bias, information, and responsiveness of bureaucrats and politicians. *Journal of Public Economics*, 97:230–244.
- Klerman, D. M. and Mahoney, P. G. (2005). The value of judicial independence: Evidence from eighteenth century England. *American Law and Economics Review*, 7(1):1.
- Kritzer, H. M. (2011). Competitiveness in state supreme court elections, 1946–2009. *Journal of Empirical Legal Studies*, 8(2):237–259.
- Kritzer, H. M. (2015). *Justices on the Ballot: Continuity and Change in State Supreme Court Elections*. Cambridge University Press.
- Landes, W. M. and Posner, R. A. (1980). Legal change, judicial behavior, and the diversity jurisdiction. *The Journal of Legal Studies*, 9(2):pp.367–386.
- Lim, C. and Snyder, J. M. (2015a). Is more information always better? party cues and candidate quality in u.s. judicial elections. *Journal of Public Economics*.

- Lim, C. H. S. (2013). Preferences and incentives of appointed and elected public officials: Evidence from state trial court judges. *American Economic Review*.
- Lim, C. H. S. and Snyder, J. M. (2015b). Is more information always better? party cues and candidate quality in u.s. judicial elections. *Journal of Public Economics*.
- Lim, C. S. H., Snyder, J. M., and Strömberg, D. (2015). The judge, the politician, and the press: Newspaper coverage and criminal sentencing across electoral systems. *American Economic Journal: Applied Economics*, 7(4):103–135.
- List, J. A. and Sturm, D. M. (2006). Elections matter: Theory and evidence from environmental policy. *Quarterly Journal of Economics*, 121(4):1249–1281. Times Cited: 5050.
- Maskin, E. and Tirole, J. (2004). The politician and the judge: Accountability in government. *The American Economic Review*, 94(4):1034–1054.
- Nelson, M. J., Caufield, R. P., and Martin, A. D. (2013). Oh, mi: A note on empirical examinations of judicial elections. *State Politics & Policy Quarterly*, page 1532440013503838.
- Pande, R. (2011). Can informed voters enforce better governance? experiments in low-income democracies. *Annual Review of Economics*, 3(1):215–237.
- Park, K. H. (2014). *Judicial Elections and Discrimination in Criminal Sentencing*. PhD thesis, Harris School, University of Chicago, Chicago, IL.
- Posner, R. (2008). *How Judges Think*. Harvard University Press.
- Posner, R. A. (1987). The law and economics movement. *The American Economic Review*, 77(2):pp.1–13.
- Pozen, D. (2010). Judicial elections as popular constitutionalism. *Columbia Law Review*, 110:2047–2134.
- Rahn, W. (1993). The role of partisan stereotypes in information-processing about political candidates. *American Journal of Political Science*, 37(2):472–496.
- Rothstein, P. (1991). Representative voter theorems. *Public Choice*, 72(2/3):193–212.
- Saporiti, A. and Tohmé, F. (2006). Single-crossing, strategic voting and the median choice rule. *Social Choice and Welfare*, 26(2):363–383.
- Shepherd, J. M. (2009). The influence of retention politics on judges’ voting. *The Journal of Legal Studies*, 38(1):169–206.
- White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*.
- Wilensky, H. L. (1964). The professionalization of everyone? *American Journal of Sociology*, 70(2):pp.137–158.

Wilson, J. (1989). *Bureaucracy: What government agencies do and why they do it*. Basic Books (AZ), New York, NY.

A Appendix

This appendix includes some further notes on the data and the institutional reforms, as well as further regression specifications.

A.1 Notes on Institutional Reforms

This section provides some notes on the institutional reforms. The key point is that there were often coterminous reforms, such as the introduction of an intermediate appellate court. To deal with this we ran all the regressions while leaving one state out. None of the results were substantially changed in these checks. Note that these coterminous reforms only threaten identification in the analysis of retention-process reforms. When we look at the electoral cycle and when we look at selection effects, we are holding court-specific incentives constant.

Colorado instituted an intermediate appellate court in 1971, four years after the election reform. Changing Colorado to a four year window does not change the results. Florida moved from partisan to nonpartisan elections in 1972, then moved from nonpartisan to merit-uncontested in 1977. In the retention-process regressions we treat these as separate reforms with five-year effect windows. Removing Florida from the regressions does not change the results.

At the same time that Illinois changed from partisan retention to uncontested retention (November 1962), the state also increased judge term lengths from nine years to ten years. However, the term-lengths change went into effect in January 1963, two years before the election reform went into effect.

At the same time it moved from partisan to merit-uncontested, Indiana increased term lengths from six years to ten years.

Kentucky instituted an intermediate appellate court at the same time that it moved from partisan to nonpartisan elections.

The Maryland governor began selecting new appointees by merit commission beginning in 1971. When it moved from nonpartisan retention to uncontested retention, the term length was reduced from 15 years to 10 years.

Oklahoma instituted an intermediate appellate court at the same time it moved from partisan to merit-uncontested.

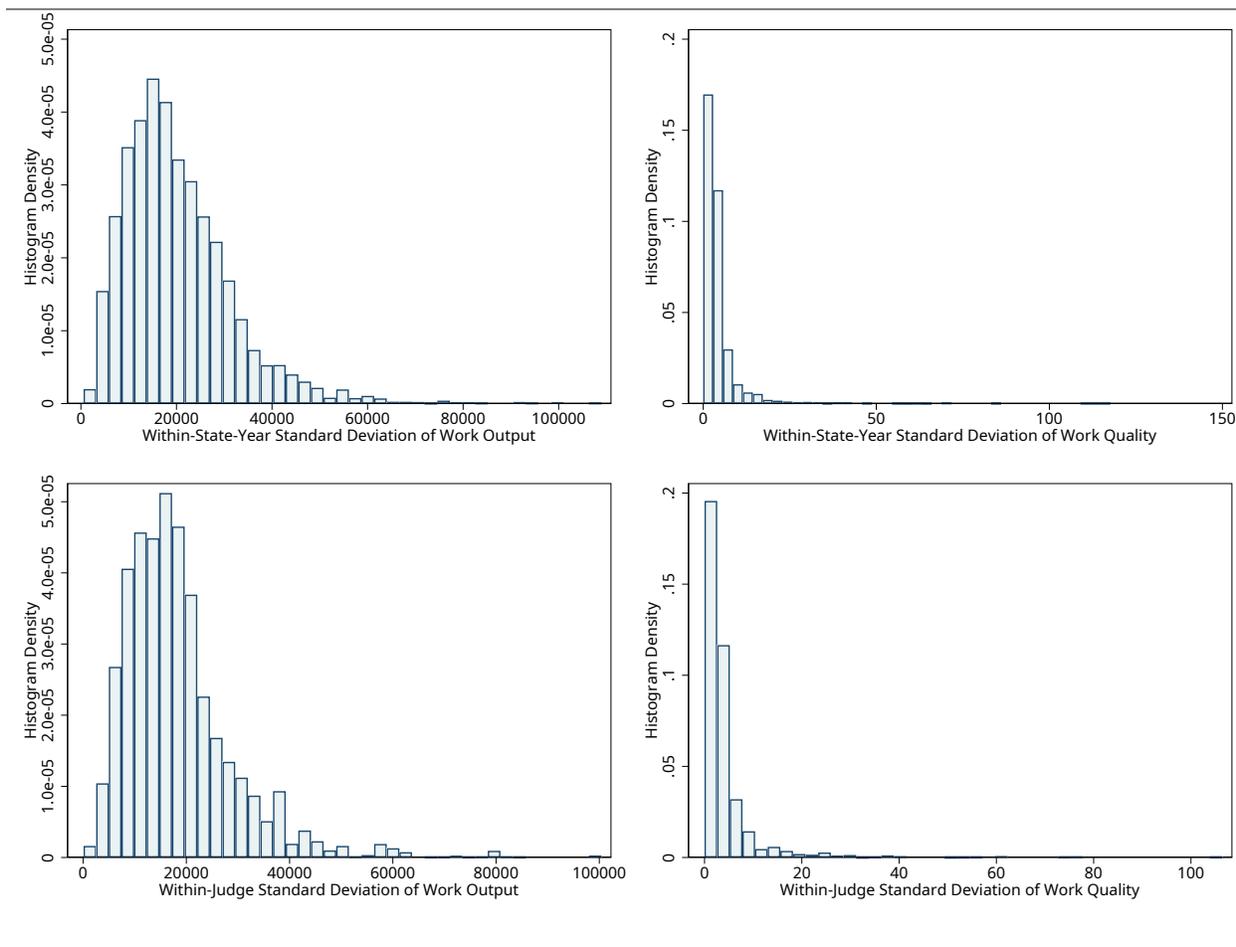
In 1973, South Dakota increased its term length from six years to eight years, eight years before the nonpartisan to merit-uncontested reform.

Figure 3: History of Selection and Retention Reforms



Notes. History of reforms to the electoral system: Partisan Elections to Non-Partisan Elections (PtoNP), Partisan Elections to Merit Selection with Uncontested Retention (PtoMU), Partisan Elections to Partisan Election Selection with Uncontested Retention, and Non-Partisan Elections to Merit Selection with Uncontested Retention (NProMU). Shaded areas indicate years included in treatment dummies. Numbers in parentheses are included when the time window used is less than ten years.

Figure 4: Histograms for Variance in Outcomes, Within-Court-Year and Within-Judge



Tennessee moved from partisan to merit-uncontested in 1972, then moved back to partisan elections in 1975. It is not included in the analysis.

Utah instituted an intermediate appellate court in 1988, two years after the reform from nonpartisan to merit-uncontested.

A.2 Appendix Tables

This section provides additional tables and empirical results.

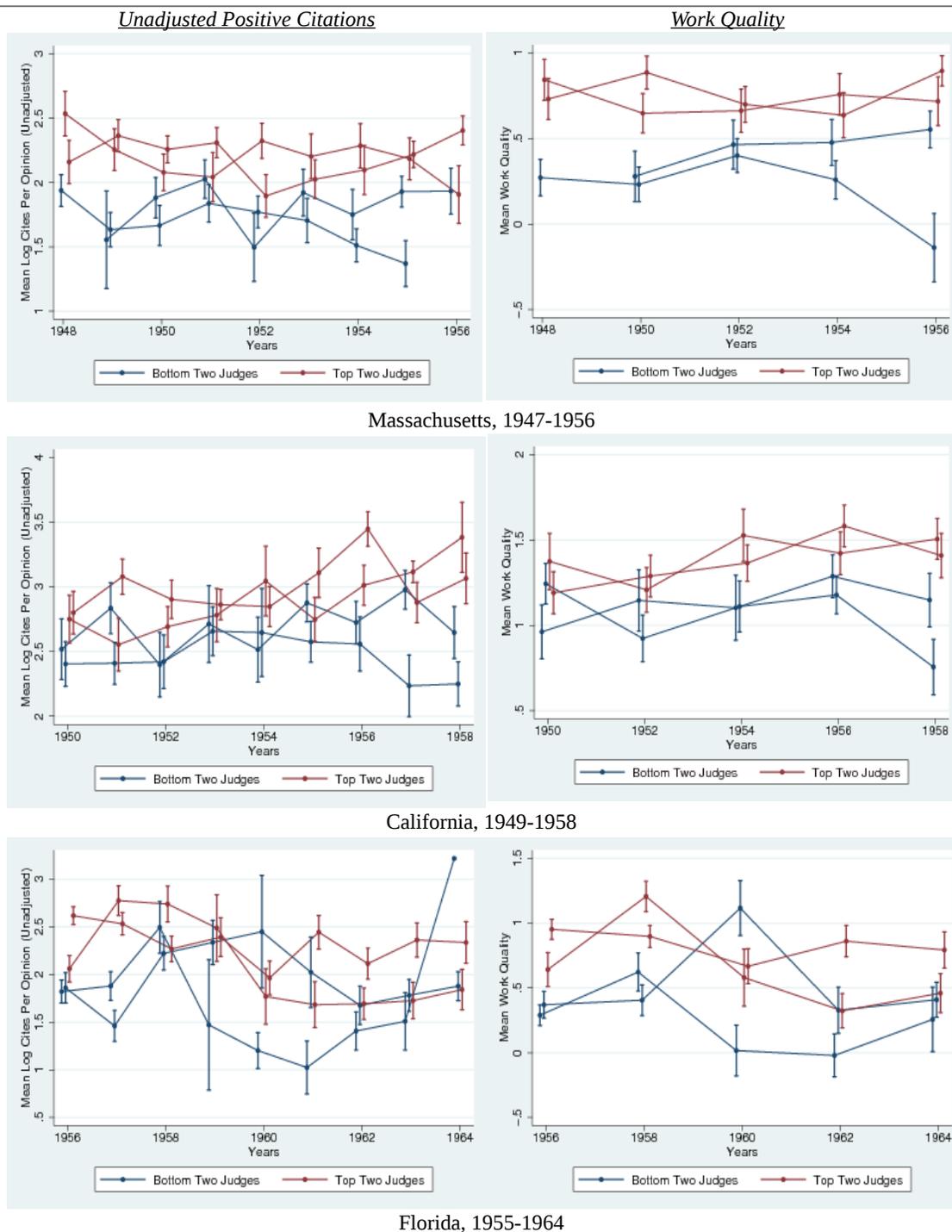
Figure 4 shows the distributions of the variances of the outcomes by court-year and by judges. This substantial variation justifies the normalization of variances within these groups of observations.

Next we look at additional outcome variables for the set of judges analyzed. Figures 7, 8,

and 9 provide judge-specific plots for three potential outcome variables to measure output. Respectively, they report the number of opinions written annually, the number of words written annually, and “work output” the number of words written per two-year period after residualizing on case characteristics. As discussed in the text, work output provides the most consistent distinctions between judges of these measures.

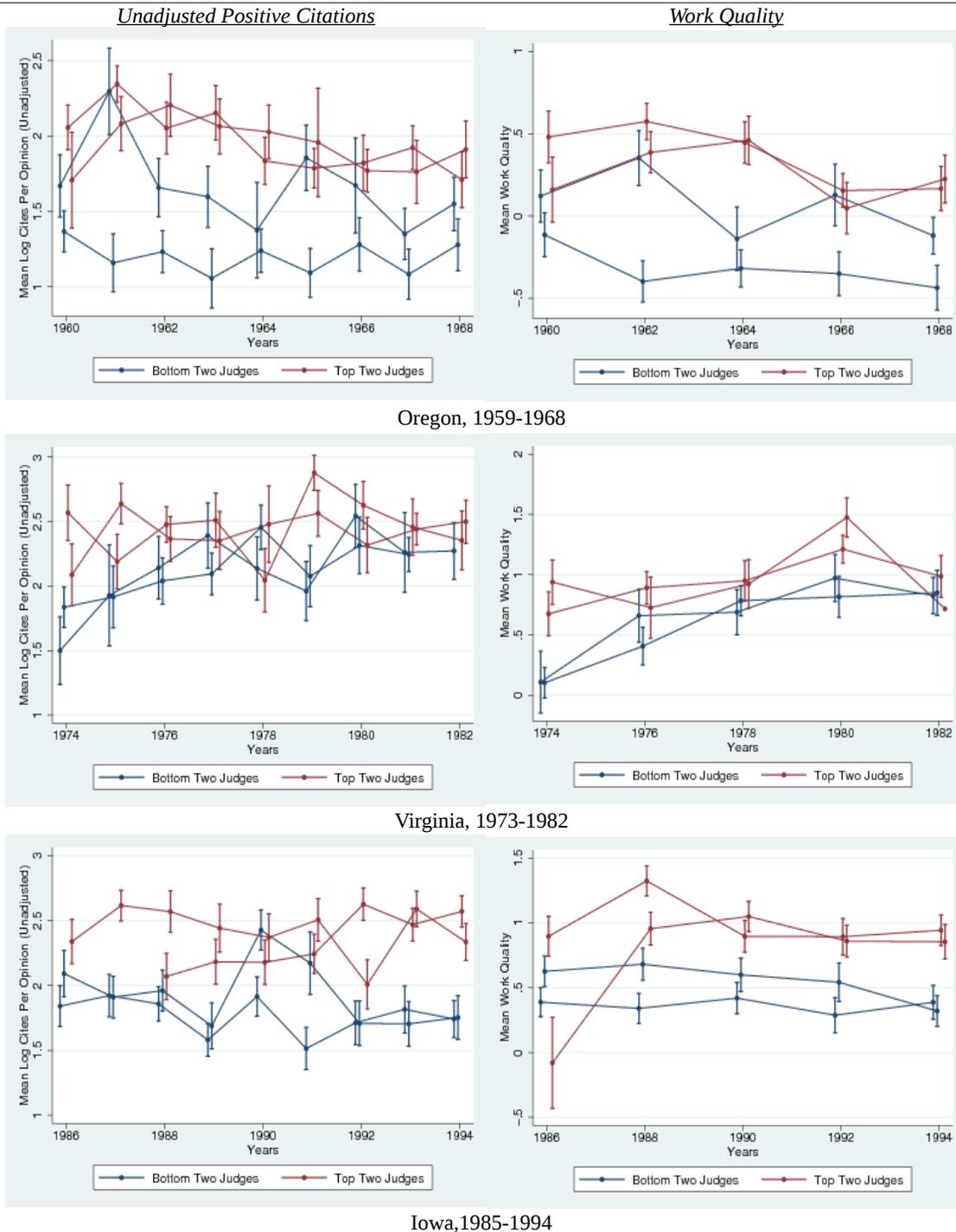
A host of other results (described briefly in the text) are shown in the replication materials.

Figure 5: Work Quality Distinctions Between Judges



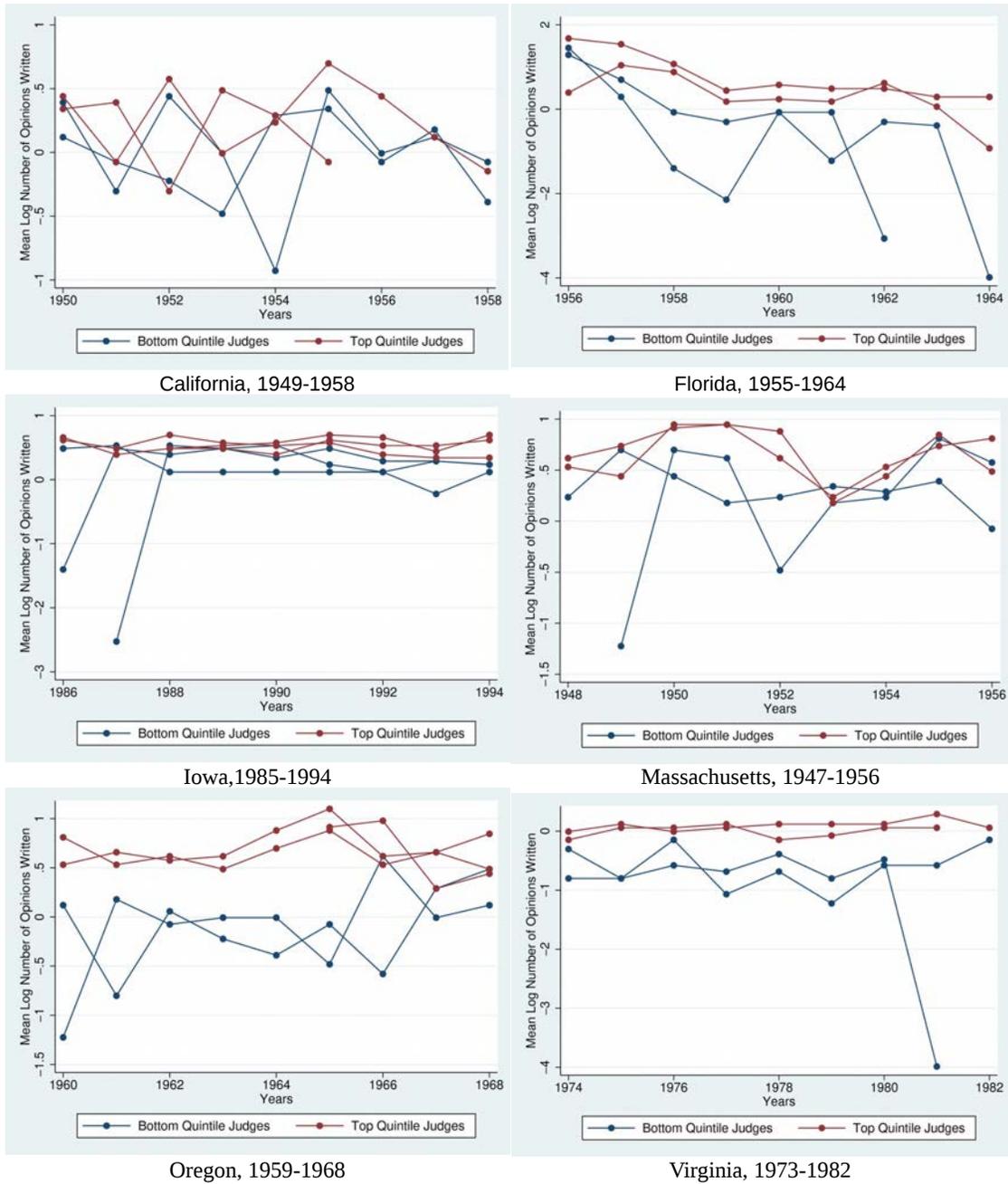
Raw data for Opinion Quality (positive citations per opinion, residualized on case characteristics and standardized) plotted by biennium, subsetted by top-quintile and bottom-quintile judges, in selection of states for periods with relatively little turnover among judges.

Figure 6: Work Quality Distinction Between Judges (cont.)



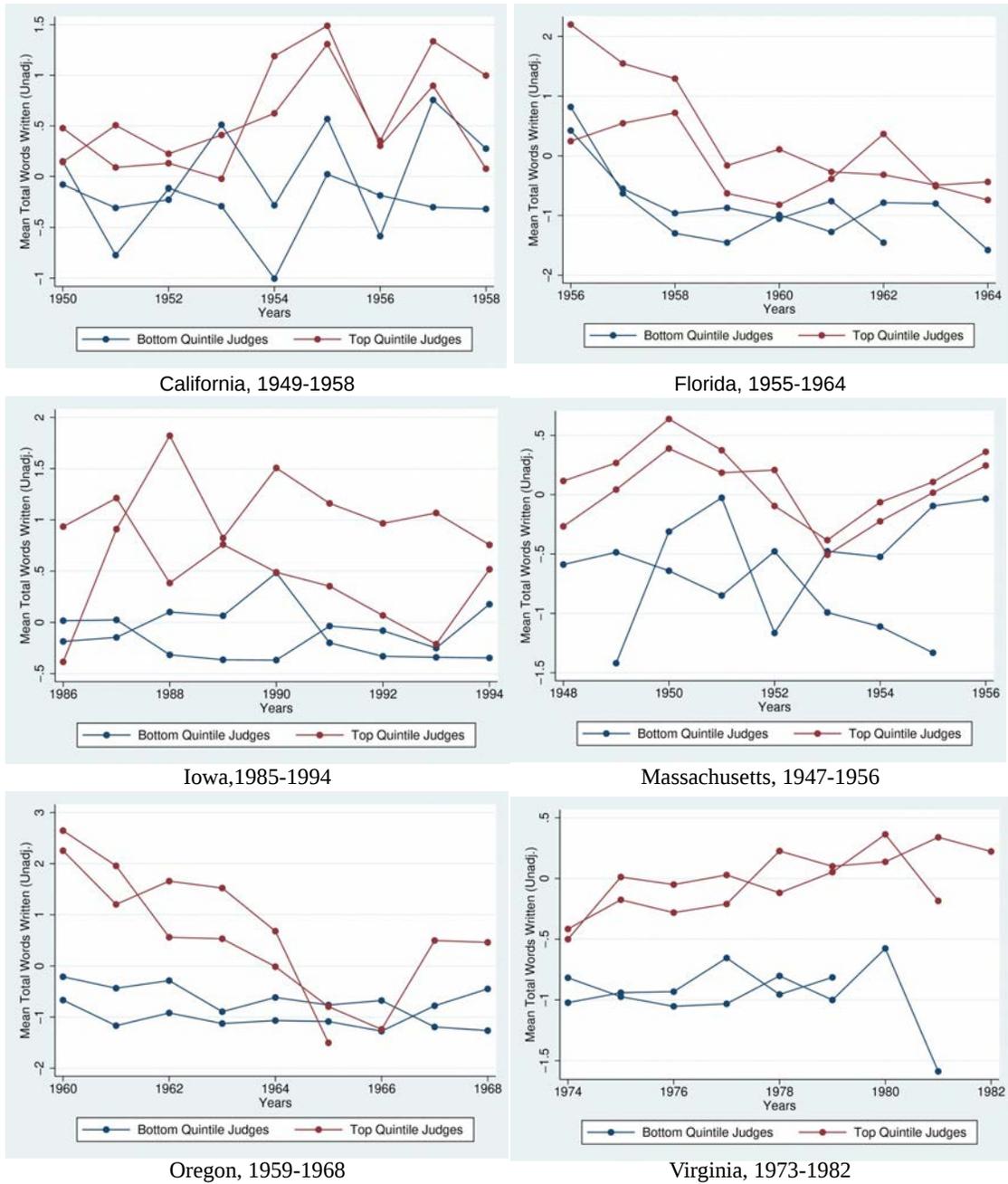
Raw data for Opinion Quality (positive citations per opinion, residualized on case characteristics and standardized) plotted by biennium, subsetted by top-quintile and bottom-quintile judges, in selection of states for periods with relatively little turnover among judges.

Figure 7: Number of Opinions Distinction Between Judges



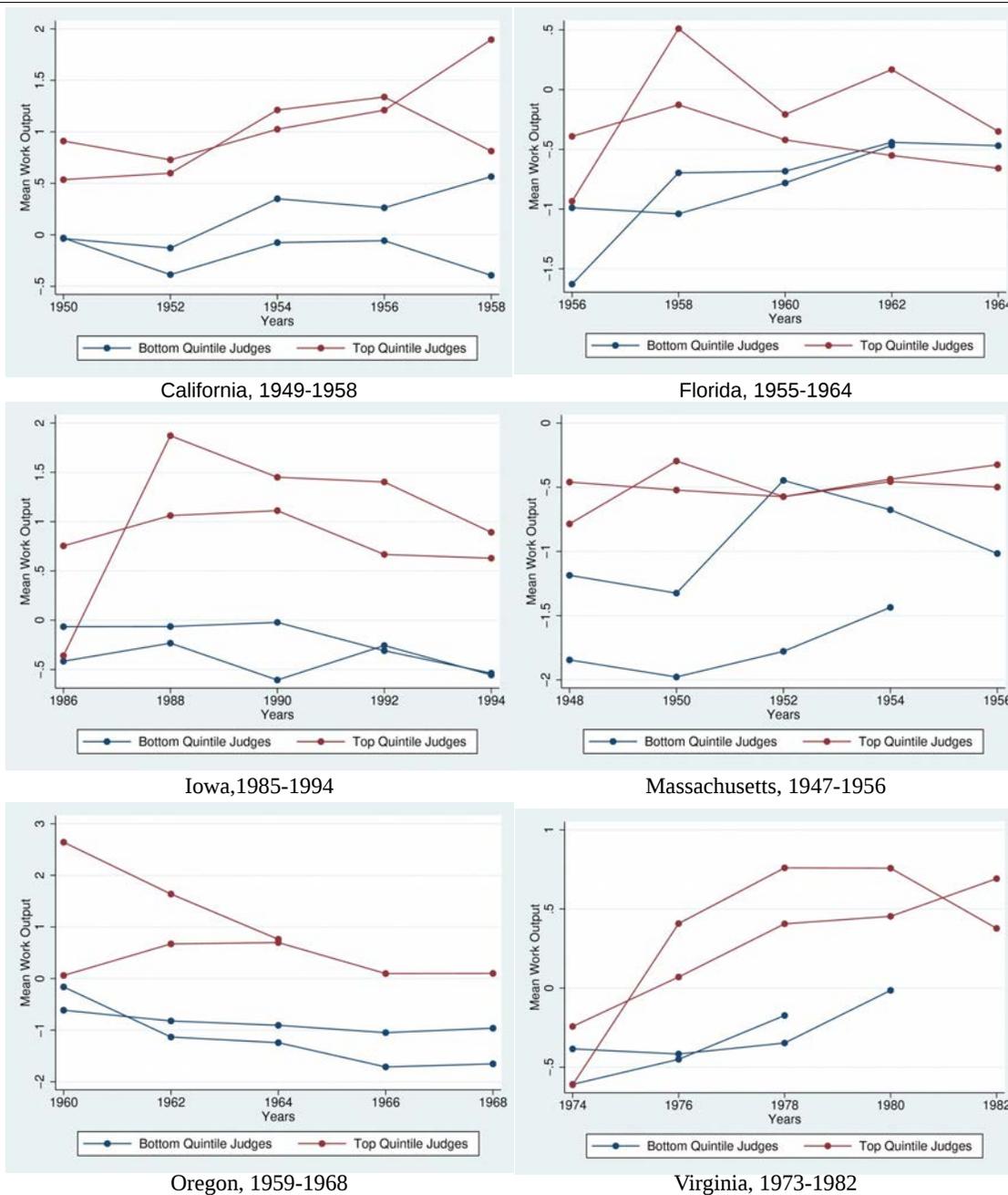
Raw data for log number of opinions plotted by year, subsetted by top-quintile and bottom-quintile judges, in selection of states for periods with relatively little turnover among judges.

Figure 8: Raw Words Written Between Judges



Raw data for number of words written plotted by year, subsetted by top-quintile and bottom-quintile judges, in selection of states for periods with relatively little turnover among judges.

Figure 9: Work Output Distinction Between Judges



Raw data for Work Output (number of words written, residualized on case characteristics and standardized) plotted by biennium, subsetted by top-quintile and bottom-quintile judges, in selection of states for periods with relatively little turnover among judges.